Journal of Hydrology: Regional Studies 6 (2016) 26–51

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Snow-melt flood frequency analysis by means of copula based 2D probability distributions for the Narew River in Poland

Bogdan Ozga-Zielinski[a], Maurycy Ciupak[b], Jan Adamowski[a,*], Bahaa Khalil[c], Julien Malard[c]

[a] Department of Environment Protection and Development, Environmental Engineering Faculty, Warsaw University of Technology, ul. Nowowiejska 20, Warsaw 00-653, Poland
[b] Hydrological Forecasting Office, Institute of Meteorology and Water Management – National Research Institute, ul. Piotra Borowego 14, Cracow 30-215, Poland
[c] Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Sainte-Anne-de-Bellevue, QC H9x3V9, Canada,

ARTICLE INFO

Article history:
Received 7 October 2015
Received in revised form 2 February 2016
Accepted 3 February 2016
Available online 24 March 2016

Keywords:
Flood frequency analysis (FFA)
2D Normal probability distribution
Copula functions
Elliptical Gaussian copula
Archimedean 1-parameter
Gumbel–Hougaard copula
Goodness-of-fit tests
Narew River
Poland

ABSTRACT

Study region: Narew River in Northeastern Poland.
Study focus: Three methods for frequency analysis of snowmelt floods were compared. Two dimensional (2D) normal distribution and copula-based 2D probability distributions were applied to statistically describe floods with two parameters (flood peak $Q_{max,f}$ and flood volume V_f). Two copula functions from different classes – the elliptical Gaussian copula and Archimedean 1-parameter Gumbel–Hougaard copula – were evaluated based on measurements.
New hydrological insights for the region: The results indicated that the 2D normal probability distribution model gives a better probabilistic description of snowmelt floods characterized by the 2-dimensional random variable ($Q_{max,f}$, V_f) compared to the elliptical Gaussian copula and Archimedean 1-parameter Gumbel–Hougaard copula models, in particular from the view point of probability of exceedance as well as complexity and time of computation. Nevertheless, the copula approach offers a new perspective in estimating the 2D probability distribution for multidimensional random variables. Results showed that the 2D model for snowmelt floods built using the Gumbel–Hougaard copula is much better than the model built using the Gaussian copula.

1. Introduction

Flood risk mapping and the attendant designation of hazard zones are an important issue facing hydrologists to ensure the safety of hydrological structures, and to protect lives, property, cultural landmarks, centres of economic activity and zones of environmental significance (Directive 2007/60/EC of the European Parliament). Resolving such questions has been largely based on an analysis of measured parameters associated with extreme flooding events (e.g. peak flow/discharge and flood timing on an annual basis). Supported by long-term measurements of these parameters, statistical methods employed

* Corresponding author.
E-mail address: jan.adamowski@mcgill.ca (J. Adamowski).

http://dx.doi.org/10.1016/j.ejrh.2016.02.001

in flood frequency analysis (FFA) allow engineers to calculate the return period (in years) of a particular maximum flood discharge ($Q_{\max,\mathrm{f}}$). This knowledge can then inform the selection of design floods for water management and, in particular, flood control structures, as well as aid in the design of flood hazard and risk zones. In turn, this can help various stakeholders manage water resources in a more effective and sustainable manner (e.g. Halbe et al., 2013; Halbe et al., 2014; Kolinjivadi et al., 2014; Straith et al., 2014; Inam et al., 2015; Butler and Adamowski, 2015).

Numerous studies have investigated the question of how to select the best probability distribution for a one dimensional (1D) random variable descriptive of peak flow during the most severe flood of a given year (Singh and Wang, 2005; Ferro and Porto, 2006; Stedinger and Griffis, 2008; Ciupak, 2013). However, in many engineering applications, the description of hydrological extreme events through a single parameter remains inadequate. When designing water management structures, it is imperative to take into account the long-term impact of peak flows on the safety, effectiveness and risk of failure of hydrological structures. This requires not only historic or predicted $Q_{\max,\mathrm{f}}$ values, but also other parameters describing the flooding event, including the related parameters of flood volume (V_f) and flood duration (T_f). To address more complex flood-related water management and water engineering issues requires the analysis of a greater number of flood parameters (Ozga-Zielinska and Brzeziński, 1997), which necessitates the use of mathematical methods capable of describing multidimensional variables. Developed by a number of investigators (Krstanovic and Singh, 1987; Yue, 1999; Zhang, 2005), the classical approach to handling these issues requires the description of various natural phenomena and their extreme events (such as floods) to employ a multidimensional normal probability distribution.

The current development of state-of-the-art computational facilities (Nourani et al., 2014) allows for the application of new approaches, including the use of probability distributions constructed with copula functions (Song and Singh, 2010; Ciupak, 2011; Jeong et al., 2013; Bačová Mitková and Halmová, 2014; Saad et al., 2014). When several probability distributions co-occur, it is necessary to develop selection criteria to assess which probability distribution best describes that of the random variable being tested.

Since the copula method considers more than one joint distribution function when estimating parameters of a multidimensional probability distribution, selecting the optimal copula function is crucial and requires one or more suitable classification criteria. These should be oriented towards assessing the goodness-of-fit between theoretical and empirical distributions in the latter's tail region, where extreme values of flood characteristics are located, and which represent conditions when the largest flooding losses occur. The large number of copula functions currently documented in the literature (Chowdhary et al., 2011; Kuchment and Demidov, 2013; Lee et al., 2013; Li et al., 2014), as well as those currently being generated, provide a significant challenge to the selection of an optimal multidimensional probability distribution by hydrologists for a given flooding situation.

Copula functions have been widely used in hydrological modelling, in particular in the flood and drought modelling fields. With regards to droughts, most research has focused on the comparison of the performance of different copula families. Sadri and Burn (2014) applied three families of Archimedian copulas (the Gumbel, Clayton and Frank copulas) for the analysis of drought severity and duration, and showed that these copulas perform similarly to bivariate approaches for short return periods but outperformed the latter in the case of longer return periods. The most appropriate copula function choice, as determined by an analysis of *Q*–*Q* plots for each, depends on the exact study site. Conversely, Lee et al. (2013) applied the Gumbel, Frank, Clayton and Gaussian copulas to bivariate (severity and duration) drought analyses in two regions of Canada and Iran and found that, while the tail-dependence of the Clayton copula was insufficient for the data set, the Frank and Gumbel copulas performed well.

In terms of flood modelling, recent research has also focused on methods for choosing the most appropriate copula. In particular, the tail dependence of various copula functions has received attention, as the models used must be well-adapted to the modelling of tail-end (e.g. extreme event) data. This was illustrated by Ganguli and Reddy (2013), who used bivariate and trivariate copulas to analyze flood peak flow, volume and duration, and found that upper tail-dependent copulas performed best for both bivariate and trivariate analyses due to the importance of extreme events in flood prediction. The objective of this study is therefore to determine whether copula functions are a more advantageous analytical tool for the probabilistic description of a 2D snowmelt flood, and the estimation of the parameters of 2D cumulative functions as compared to the 2D normal distribution with parameters estimated by the ML method.

The Archimedean Gumbel–Hougaard copula, in particular, has performed well in past analyses. Karmakar and Simonovic (2009) report that this copula was the most appropriate for bivariate flood peak flow-volume analysis in the Red River (North Dakota, USA). Meanwhile, Poulin et al. (2007), in their comparison of the application of seven copula families to flow data of the Loire River (France), showed that copulas with upper tail independence (such as the Clayton and Frank copulas) dramatically over-estimated return periods of floods, while the Gumbel and survival Clayton copulas performed well. A different approach to determining the optimum choice of copula function was proposed by Domino et al. (2014), who suggested a new methodology for copula analysis in connection with the formalism of Detrended Fluctuations Analysis (DFA) and Anomalous Diffusion (AD) for the prediction of negative and positive auto-correlations. The theory was used for the statistical analyses of weakly predictable maximum storm tides recorded at five different harbours in the Baltic Sea. The authors detected negative and positive auto-correlations which can be understood as the low or high probability for the next extreme storm tide, which follows the previous extreme event.

In a very exhaustive application of various copula options, Sraj et al. (2014) applied Archimedian (Ali–Mikhail–Haq, Clayton, Frank, Gumbel–Hougaard, and Joe), elliptical (e.g. Normal and Student-*t*) and extreme value (*e.g.* Galambos, Hüsler–Reiss and Tawn) copulas to the bivariate analysis of peak discharge-volume, peak discharge-duration, and volume–duration variable pairs from 58 flood events on the Sava River (Slovenia). They determined that the Gumbel–Hougaard copula was the most appropriate for peak discharge–volume bivariate analyses. The results of Zhang and Singh (2006) concur, as the Archimedean copulas (Gumbel–Hougaard, Ali–Mikhail–Haq, Frank, and Cook–Johnson) they applied to bivariate flood peak flow-volume analysis, and the Gumbel–Hougaard copula in particular, outperformed normal distribution modelling of the variables.

Conversely, other studies have identified other copulas as the most appropriate for flood data analysis in their particular cases. For instance, Renard and Lang (2007) successfully applied the Gaussian copula to flood data for three different regions; although they caution that the Gaussian copula is unsuitable in cases where tail independence cannot be demonstrated. Wong et al. (2010) used trivariate Gumbel–Hougaard and *t*-copulas to analyze flood peak and average intensity as well as duration, and found that both gave similar results but concluded that, although the Gumbel–Hougaard copula is easier to fit, it does impose more restrictions on outer correlations than the *t*-copula. Zhang and Singh (2007), in their application of four Archimedean copulas (Gumbel–Hougaard, Ali–Mikhail–Haq, Frank, and Cook–Johnson) to rainfall bivariate analysis, showed that only the Frank copula was suitable for the analysis of both highly negatively and positively correlated variables. Similarly, Favre et al. (2004) applied the Clayton and Frank copulas to bivariate analysis of flood data in the Rimouski River (Québec, Canada) and showed slightly better behavior for the Frank copula despite the fact that Chowdhary et al. (2011) successfully applied the Clayton copula for peak flow and volume bivariate analysis.

In recent years, a rapid increase in copula functions has occurred; a significant number of parametric bivariate copulas now exist. While one could assume that the choice of a particular copula might depend on the nature of the particular data set, it is also important to consider the results of the application of different copula functions to flood events and the practical significance of differences in the model outputs. Consequently, in an attempt to better describe quantitative characteristics of extreme snowmelt floods, this study compared two-dimensional (2D) normal probability distributions to distributions developed with selected 2D copula functions. This was done to address the concern of whether the use of the classical approach in the form of a 2D normal probability distribution – in comparison to the more sophisticated copula methods – is a sufficiently accurate description of ($Q_{\max,\mathrm{f}}$, V_f) from the point of view of engineering applications.

Data preprocessing in this study included the identification of snowmelt floods and an analysis of the homogeneity of data, leading to a probability analysis of quantitative characteristics of snowmelt flood extremes. The present analysis of snowmelt floods compared a 2D normal probability distribution with parameters estimated by a Maximum Likelihood Method to two statistical models with parameters estimated by different copula methods. The snowmelt floods were characterized by the 2-dimensional random variable, ($Q_{\max,\mathrm{f}}$, V_f): the maximum flood discharge and total volume. The random variable sample was constructed on the basis of floods from 1966 to 2012, with observations from the Wizna hydrological monitoring station in the Narew River watershed in Poland.

This analysis included: (i) an assessment of marginal probability distributions' statistical properties, (ii) a selection of the best fitted marginal probability distributions, (iii) the estimation of density function parameters for the 2-dimensional probability distribution, (iv) the realization of a 2D random variable, and (v) assessing the goodness-of-fit between the theoretical and empirical distributions. The hydrological analysis was supported by sophisticated statistical and graphical procedures. A theoretical description of the density function of a 2D probability distribution, estimated by either the maximum likelihood method (MLM) or a selected copula function, is followed by a presentation of the analytical procedures of data processing as well as the qualitative and quantitative comparison of probability distributions.

2. Theoretical background

The snowmelt flood frequency analysis for the Narew River in northeastern Poland employed historical measurements of peak discharge [$Q_{\max,\mathrm{f}}$ ($\mathrm{m^3\,s^{-1}}$)] and flood volume [V_f ($10^6\,\mathrm{m}^3$)] at the Wizna station (Fig. 1) for the period of 1966–2012.

Over the past few decades, there has been an increase in the nonstationarity of random variables describing extreme natural phenomena (such as snowmelt floods) (Belayneh et al., 2014). Such disturbances may be of natural and anthropogenic origin. While the first type of disturbance is independent of human activity, the anthropogenic factors arise directly from their operations in the valley of the river and catchment area.

Nonstationarity of the random variable affects defining the return period T and thus the probability exceedance p, i.e., for stationary univariate analysis

$$T = \frac{\mu}{1 - F(x)} \tag{1}$$

where: x is the realization of variable X, $\mu > 0$ denotes the average inter-arrival time between two realizations of the process, $F(x) = 1\text{-}p$ indicates the distribution function of X and for nonstationary univariate analysis

$$T = \frac{1}{\sum_{j=1}^{T} p_j} = \frac{1}{\bar{p}} = \frac{1}{1 - \overline{F_j(x)}} \tag{2}$$

where: p_j changes at each time step j along time series.

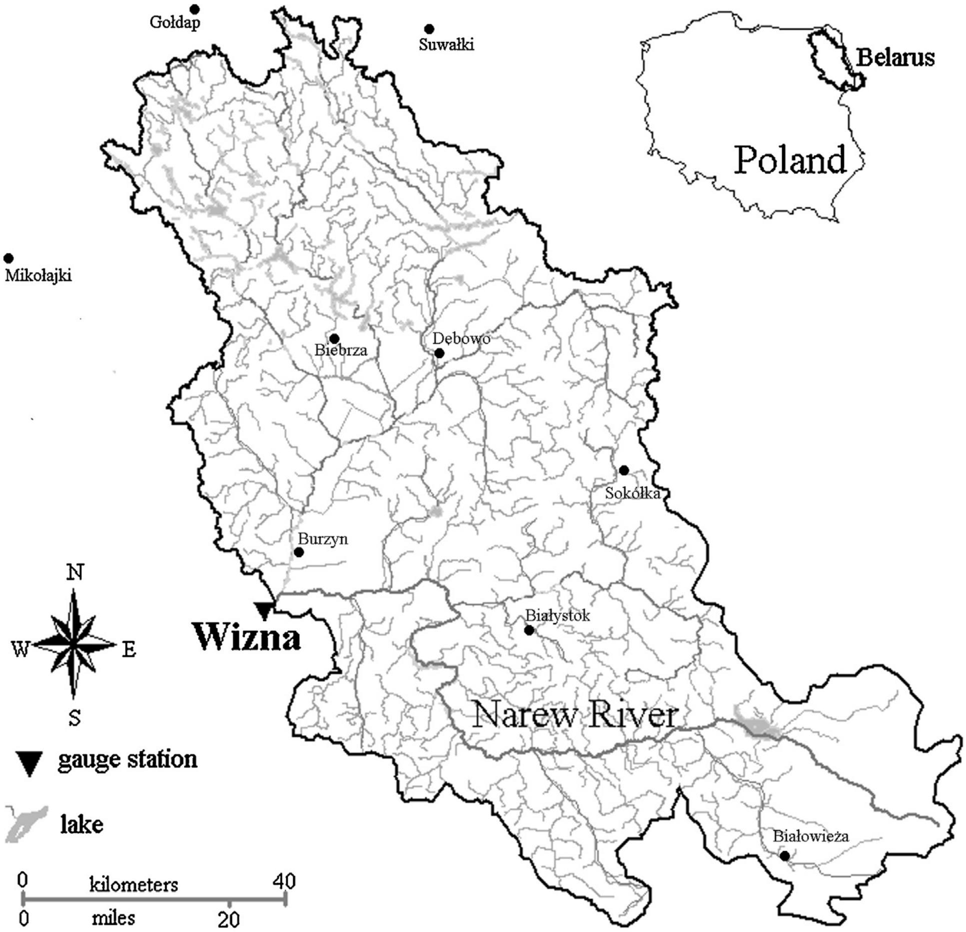

Fig. 1. Hydrographic map of the Upper Narew River watershed, Poland, including the location of the Wizna hydrological station situated at the watershed outlet.

Hence, it follows that the form of T and thus p depends on the hypothesis regarding the stationarity or nonstationarity of the univariate random variable analyzed. From this point of view, in this study we used a strong analysis to examine the homogeneity of random variables described in Section 3.2 and a statistical analysis of the quantitative characteristics of snowmelt floods in Section 3.3, especially in the context of their independence and stationarity. In addition, it must be noted that the use of T does not result in additional information in comparison to p, therefore in this study we used the concept of probability exceedance.

The problem becomes more complicated with the transition from univariate analysis to the case of multivariate analysis. This study refers to the case of 2D analyses. Snowmelt floods were characterized using 2D variables ($Q_{max,f}$, V_f). The use of copula functions to estimate the parameters of the 2D distribution function made it easier to incorporate different cases involving stationary and/or nonstationary phenomena.

Knowledge regarding the form of the exceedance probability function for the ($Q_{max,f}$, V_f) variable is essential in the analysis of long-term impact of high water on water safety structures. This is analyzed in the context of snowmelt flood occurrence, and water management structure design, as well as the assessment of their effectiveness and risk of their future failure. By definition, a copula is a joint cumulative distribution of many random variables such that each yields a uniform distribution on the segment [0,1] (Sklar, 1959; Nelsen, 2006). For two variables, the copula function C adopts the following general formula:

$$C(v_1, v_2) = P(V_1 \leq v_1, V_2 \leq v_2) \tag{3}$$

The functional form of the copula function does not determine the distribution of the marginals and merely determines the dependence between the two random variables, yet has no influence on the marginals themselves. The random variables V_1 and V_2 are cumulative distributions of random variables X_1 and X_2, i.e $V_1 = F_1(X_1)$, $v_1 = F_1(x_1)$ and $V_2 = F_2(X_2)$, $v_2 = F_2(x_2)$, where $X_1 = Q_{max,f}$ and $X_2 = V_f$. In the case of 2D analysis, it is relatively easy to describe the following probabilities via copula functions (Cherubini et al., 2004; Klein et al., 2011):

$$P(V_1 \le v_1, V_2 > v_2) = v_1 - C(v_1, v_2) \tag{4}$$

$$P(V_1 > v_1, V_2 \le v_2) = v_2 - C(v_1, v_2) \tag{5}$$

$$P(V_1 \le v_1 | V_2 \le v_2) = \frac{C(v_1, v_2)}{v_2} \tag{6}$$

$$P(V_1 \le v_1 | V_2 > v_2) = \frac{v_1 - C(v_1, v_2)}{1 - v_2} \tag{7}$$

$$P(V_1 \le v_1 | V_2 = v_2) = \frac{\partial C(v_1, v_2)}{\partial v_2} \tag{8}$$

$$P(V_2 \le v_2 | V_1 = v_1) = \frac{\partial C(v_1, v_2)}{\partial v_1} \tag{9}$$

and, exceedance probability exceeding V_1 and V_2 (called the survival 2-copula of V_1 and V_2) which can be used to calculate the joint survival function $F_{V_1 V_2}(v_1, v_2)$ as

$$P(V_1 > v_1 \wedge V_2 > v_2) = 1 - v_1 - v_2 + C(v_1, v_2) \tag{10}$$

and exceedance probability exceeding V_1 or V_2

$$P(V_1 > v_1 \vee V_2 > v_2) = 1 - C(v_1, v_2) \tag{11}$$

Making comparisons between the above described probabilities, which are defined in different domains, and over different sets and subsets of data is not exactly correct, because it introduces misconceptions in the interpretation of the results. In this study, the extreme events associated with the occurrence of snowmelt floods in their extremely unfavorable conditions were analyzed; therefore the exceedance probability exceeding $Q_{max,f}$ and V_f (Eq. (10)) was selected.

Floods were described probabilistically using either a 2D normal probability distribution or 2D distributions built with one of two copula functions. In the first case, the assumption was made that if the marginal distributions of $Q_{max,f}$ and V_f were normal then the probability distribution of the 2-dimensional variable ($Q_{max,f}$, V_f) would also be normally distributed (Kotz et al., 2000; Myung, 2003).

In the second case, two copulas were used: elliptical Gaussian (Genest et al., 2007; Joe, 1997; Renard and Lang, 2007) and Archimedean 1-parameter Gumbel–Hougaard (Gumbel, 1960; Salvadori and De Michele, 2007). Parameter estimation occurred in two stages: (i) estimation of marginal probability distribution parameters by the MLM method, and (ii) estimation of the correlation coefficient (ρ) through Bayesian simulation or Inference Functions for the Margins (IFM) method for elliptical Gaussian (Danaher and Smith, 2009) and Archimedean 1-parameter Gumbel–Hougaard (McLeish and Small, 1988; Choroś et al., 2010) copulas, respectively.

2.1. 2D density function for a normal distribution and estimation of its parameters by MLM

The random variable (X_1, X_2) was assumed to yield a normal probability distribution. However, separate variables X_1 and X_2 can yield various probability distributions. Therefore, in such cases, the variables X_1 and X_2 require a transformation to a normal distribution. The formula $u = \ln(x)$ was used as a normalization function. The variables X_1 and X_2 were denoted as U_1 and U_2 respectively after normalization.

The density function of a 2D normal distribution can be formulated as (Rose and Smith, 2002):

$$f(u_1, u_2) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left[\frac{-(\boldsymbol{u} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u} - \boldsymbol{\mu})}{2}\right] \tag{12}$$

where,

$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ is the matrix of values for two quantitative characteristics of the flood;

$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ is the matrix of mean values of two quantitative characteristics of the flood;

$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{bmatrix}$ is the variance-covariance matrix;

$\boldsymbol{\Sigma}^{-1}$ is the inverse of $\boldsymbol{\Sigma}$;

$|\boldsymbol{\Sigma}|$ is the determinant of matrix $\boldsymbol{\Sigma}$, and

$(\mathbf{u}\text{-}\boldsymbol{\mu})^T$ is the transpositive matrix.

The MLM method was used to estimate 2D normal distribution parameters. The estimator of the mean value ($\hat{\boldsymbol{\mu}}$) was calculated as:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{u}_i \tag{13}$$

where,

N is the size of the measurement series (random variable sample).

The unbiased estimator of the covariance matrix ($\hat{\Sigma}$) was calculated as:

$$\hat{\sum} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{u}_i - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{u}_i - \hat{\boldsymbol{\mu}}\right)^T \tag{14}$$

2.2. General form of copula function for a 2D distribution

The copula theory can be employed to build a 2-dimensional probability distribution for any marginal distribution (De Michele and Salvadori, 2003; Salvadori et al., 2007). The random variable (X_1, X_2) can be described through a 2D Gaussian distribution, whereas variable X_1 may yield a log-normal marginal distribution while variable X_2 yields a Gumbel marginal distribution. The copula theory allows one to build 2D distributions for any marginal distributions.

Given two random variables X_1 and X_2 with cumulative distribution functions (CDF) $F_1(X_1)$, $F_2(X_2)$ the goal of using copula theory is to obtain a 2D distribution function in the form of $F(X_1, X_2)$. Sklar (1959) proved that a C function (copula) exists as follows:

$$F(X_1 = x_1, X_2 = x_2) = C[F_1(x_1), F_2(x_2)], x_1, x_2 \in R \tag{15}$$

The function C is a CDF joining marginal distributions $F_1(X_1)$, $F_2(X_2)$.

For a 2D case, $v \in [0,1]^2$, the function $C(v)$ is equal to zero if at least one element of v is equal to zero, e.g. $C(v_1 = 0, v_2) = 0$ or $C(v_1, v_2 = 0) = 0$. Whereas $C(v) = v_j$ for $j = 1, 2$, if all coordinates of v are equal to 1 except v_j. For a 2D case, the above relationship occurs when $C(v_1 = 1, v_2) = v_2$ or $C(v_1, v_2 = 1) = v_1$. In the terminology of copulas, these properties can be presented as $C(F_1(x_1), 1) = F_1(x_1)$, $C(1, F_2(x_2)) = F_2(x_2)$. This means that for any X_1, X_2 variables the cumulative distribution function F constructed using copula function C will always have desired marginals F_1 and F_2. Thus, by drawing from the F, the X_1, X_2 will retain their distributions. In this study, copulas are used as a testing method for unscaled measures of dependence and to construct a 2D distribution of the desired properties as well as to simulate the 2D variable. Thus, the role of the joint function C is to model the relationship between variables X_1, X_2. In the case of marginal distributions of continuous variables, by differentiating Eq. (15), the 2-dimensional density function can be obtained:

$$f(X_1 = x_1, X_2 = x_2) = c[F_1(x_1), F_2(x_2)]f_1(x_1)f_2(x_2) \tag{16}$$

where:

$c[F_1(x_1), F_2(x_2)]$ is the density of the copula and can be obtained by computing the second partial derivative of the copula on the two variables (arguments):

$$c\left[F_1(x_1), F_2(x_2)\right] = \frac{\partial^2 C}{\partial v_1 \partial v_2} \text{and } v_1 = F_1(x_1), v_2 = F_2(x_2) \tag{17}$$

Eq. (16) illustrates how the density function c takes into account the relationship between variables X_1, X_2. For example, if $C(v_1, v_2) = v_1 v_2$, then $c(v_1, v_2) = 1$. The 2D function $c(v_1, v_2)$ is now the product of its marginal distributions in such a manner that the random variables X_1 and X_2 are independent. The copula $C(v_1, v_2) = v_1 v_2$ is the simplest copula function concerning independent random variables.

2.3. Justification for the choice of copula functions for 2D probability analysis

The selection process of copula functions in this study was performed intuitively (Chowdhary et al., 2011). This process resulted from designated acceptable dependence ranges and tail dependence characteristics of the data under consideration. In this study, the classical approach in the form of a 2D normal distribution is compared with an elliptical Gaussian and an Archimedean 1 parameter Gumbel–Hougaard copula function. The primary argument for the choice of the Gaussian copula stems from its construction through a 2D normal distribution over R^2 by using the probability integral transformation, whereas Archimedean copulas are an associative class of copulas and represent very different copula families. The Archimedean class of copulas is commonly used and includes a whole suite of closed-form copulas that cover a wide range of dependency structures (i.e. a range of correlation (nonlinear) frequencies observed in hydrology), including comprehensive and non-comprehensive copulas, radial symmetry and asymmetry, and asymptotic tail dependence and independence (Chowdhary et al., 2011). Most common Archimedean copulas offer an explicit formula, which is otherwise not possible with the Gaussian copula. In practice, Archimedean copulas are popular because they allow modelling dependence in arbitrarily high dimensions with only one parameter in governing the strength of dependence. In the case of a Gaussian copula with increasing dimensions, the number of parameters needed for model calibration increases dramatically (i.e. increasing the size of the variance–covariance matrix).

Genest and Favre (2007) tested some extreme-value copulas and suggested that copulas such as the Gumbel–Hougaard are good tools for bivariate modelling of the hydrological pair ($Q_{\max,f}$, V_f). The Gumbel–Hougaard, as an extreme-value copula, not only arises naturally in the domain of extreme-value theory but can also be a convenient choice to model general positive dependence structures. Calculation of the probability of a flood exceeding a certain threshold requires knowledge of the joint distribution of maxima flows and flood volumes during the forecasting period. This is a typical field of application for extreme-value theory, in our case for the Gumbel–Hougaard copula. In such situations, extreme-value copulas can be considered to provide appropriate models for the dependence structure between exceptional events.

2.4. Form of elliptical Gaussian copula and Archimedean Gumbel–Hougaard copula for the 2D distribution of a random variable ($Q_{max,f}$, V_f)

The Gaussian copula can be defined as the double integral (Cherubini et al., 2004; Garcia and Gençay, 2007):

$$C_\rho^G(v,z)=\Phi_\rho[\Phi^{-1}(v),\Phi^{-1}(z)] \tag{18}$$

in which,

$$\Phi_\rho(\Phi^{-1}(v),\Phi^{-1}(z))=\int_{-\infty}^{\Phi^{-1}(v)}\int_{-\infty}^{\Phi^{-1}(z)}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[\frac{2\rho su-s^2-u^2}{2(1-\rho^2)}\right]dsdu \tag{19}$$

where, in the present case, $v=F_1$ ($Q_{\max,f}$) and $z=F_2(V_f)$, and the random variables $Q_{\max,f}$ and V_f can yield any probability distribution.

A major point of difference among possible copula functions is the range of correlation coefficients. The Gaussian copula has nearly the full $(-1, 1)$ range in pair-wise correlation and is therefore a general and robust copula for most applications. Furthermore, the Gaussian copula has a desirable property in that, as the number of dimensions (m) increases, the number of parameters in the multivariate density increases only in the order of m^2 (Danaher and Smith, 2009). In the case of high-dimension distributions, the estimation of parameters using the traditional maximum likelihood method can be infeasible for the Gaussian copula; therefore, in this study, the MCMC simulation algorithm was applied (Andrieu et al., 2003).

The Gumbel–Hougaard copula (Gumbel, 1960; Hougaard, 1986, 2000), used for extreme values of a random variable, is given as:

$$C_\theta^{GH}(v,z)=\exp\left\{-[(-\log v)^\theta+(-\log z)^\theta]^{\frac{1}{\theta}}\right\} \tag{20}$$

where: $\theta\in[1;+\infty)$.

A Gumbel–Hougaard copula is considered an extreme value copula since it is a copula C such that $C(v,z)=\lim_{N\to\infty}C^N\left(v^{1/N},z^{1/N}\right)$. C function implies a copula representing a set of independent and identically distributed random variables ($Q_{\max,f,i}$, $V_{f,i}$) where $i=1\ldots N$, and C is the joint distribution of their component-wise maxima $Q_{\max,f}(N)$ and $V_f(N)$ (Chowdhary et al., 2011). The Gumbel–Hougaard copula is also an Archimedean copula along with being an extreme value copula, since it can be written in the form of an Archimedean copula with a generator function, i.e. $\phi[F(x_1,x_2)]=\phi\left\{C[F_1(x_1),F_2(x_2)]\right\}=\phi[F_1(x_1)]+\phi[F_2(x_2)]$, where ϕ is called a generator of copula. Irrespective of the distribution dimension, the model structure is described by a single parameter θ. This parameter measures the degree of dependence ranging from independence ($\theta=1$) to complete dependence ($\theta=+\infty$). For the Gumbel–Hougaard copula, extension to negative dependence is not possible, thus this copula can represent independence and positive dependence only (Genest et al., 2007). The estimation of this parameter was carried out using the IFM method. This distribution can be generated via mixtures of certain extreme-value distributions over stable distributions, a representation that yields large possibilities for modelling a phenomenon with a complex structure (e.g. snowmelt flood).

3. Methodology

To compare the two methods of parameter estimation for 2D probability distributions, the following steps were followed:

1. Identification of snowmelt floods (based on Ciupak, 2004; Jeong et al., 2013).
2. Analysis of flood homogeneity $Q_{\max,f}$ and V_f series (based on Ozga-Zielinski, 1999; Ozga-Zielinska et al., 2005).
3. Statistical analysis of random variables $Q_{\max,f}$ and V_f, including (Ciupak, 2004):

a. normalization and verification (Strupczewski, 1967)
b. calculation of statistical parameters of the random variables $Q_{max,f}$ and V_f (Węglarczyk, 2010);
c. consideration of mathematical models of probabilistic properties of $Q_{max,f}$ and V_f (Brzeziński, 2010; Jeong et al., 2013), and
d. estimation of parameters of marginal probability distribution functions.

4. Selection of the best fitted marginal distribution function for $Q_{max,f}$ and V_f (Koziol, 2008; Genest et al., 2009).
5. Estimation of parameters of 2D normal distribution function and generation of 2D random variable $(q_{max,f}, v_f)$ (Klonecki, 1999; Kotz et al., 2000).
6. Estimation of parameters of the function of the 2D copula and generation of 2D random variable $(q_{max,f}, v_f) = (F_1(q_{max,f}), F_2(v_f))$ including:
   a. estimation of elliptical Gaussian copula (Danaher and Smith, 2009), and
   b. estimation of Archimedean 1-parameter Gumbel–Hougaard copula (Cherubini et al., 2004).
7. Goodness-of-fit measures for 2D probability distributions (Kotz et al., 2001; Genest et al., 2006, 2009; Koziol, 2008).

3.1. Identification of snowmelt floods

A total of 44 snowmelt flood events were identified through a series of river discharge measurements (1966–2012) obtained at the Wiza gauging station on the Narew River in northeastern Poland. A flood was defined as a period when discharges equalled or exceeded a threshold determined according to the hydrological criterion of $Q^{H}_{threshold} = Q_{max}^{1966-2012/Nov-Apr} = 73.0\ m^3 s^{-1}$(Ozga-Zielinska and Brzeziński, 1997). To specifically identify snowmelt floods, the hydrological stage of the catchment prior to the occurrence of the flood, as well as the hydro-meteorological factors affecting floods in September and October every year between 1966 and 2012, were taken into account. In doing so, potential floods from November, December and January arising from non-standard fall and early spring rainfall events were eliminated. This analysis was carried out to obtain an *a priori* genetic homogeneous series of floods that could serve as the base for statistical analysis and the computation of quantitative characteristics for $Q_{max,f}$ and V_f.

The analysis of genetic (physical) conditions of flood occurrence take into account the meteorological variables of: daily and monthly precipitation totals, type of precipitation (rain, snow, and mixed rain and snow fall), snow cover depth, water equivalent of snow, mean daily air temperature, minimum air temperature on the ground, daily temperature distribution, and state of soil (daily mean soil temperature at 5 cm depth). The main sources of meteorological information were measurements and observations carried out at meteorological stations in Bialystok, Mikolajki and Suwalki, climatic stations in the Biebrza, Goldap and Bialowieza regions, respectively, and precipitation recording sites in Burzyn, Debowo and Sokolka (Fig. 1).

3.2. Homogeneity analysis

The 1D and 2D probability distributions used in this paper can only be employed with homogeneous time series; however, hydrological data series can include significant non-homogeneities due to human interventions, changes in measurement equipment or methods, and natural changes in the system. We therefore conducted a homogeneity analysis of the data according to the procedure proposed by Ozga-Zielinski (1999) and Ozga-Zielinska et al. (2005). This includes graphical analysis and correction of non-homogeneities due to changes in the basin or measurement instruments (e.g. genetic method) as well as statistical methods of checking for outliers (e.g. Grubbs–Beck test), independence (e.g. Wald–Wolfowitz test and Anderson serial correlation coefficient test) and stationarity of the data (e.g. Kruskal–Wallis test for jump of mean value, Spearman rank correlation (SRC) coefficient tests for the trend of mean value, and the SRC for trend of variance). Untransformed data series were used for the homogeneity analysis, as variable normalization could mask potential non-homogeneities.

3.3. Statistical analysis of flood quantitative characteristics $Q_{max,f}$ and V_f

Input data were processed separately for the estimation of 2D normal probability distributions of random variables $(Q_{max,f}, V_f)$ by the MLM method and for estimation of 2D distributions using copula functions. In the first case, the normality of marginal distributions is required. The verification of goodness-of-fit, theoretical and empirical distributions employed the λ–Kolmogorov test at a significance level of $\alpha = 0.05$, with a reduced critical value to account for the fact that parameters of the normal distribution were estimated rather than being known *a priori* (Węglarczyk, 1993). According to Węglarczyk (1993), the percent of reductions in critical values $\lambda_{cr}(\alpha)$ for generating quantiles for $\alpha = 0.10$, 0.05 and 0.01, are 33, 35 and 37% respectively for asymptotic values. For further calculations, the critical value $\lambda_{cr}(\alpha)$ was reduced 35% for a significance level of $\alpha = 5\%$ (0.05).

The goodness-of-fit analysis was also applied to Normal Probability Plots. When empirical and normal distributions diverged, a normalization was applied such that $u = \ln(x - c)$, where c was the lower bound of the probability distribution and u was the normalized variable. The main reasons for divergences are anthropogenic factors (i.e. factors related to the direct and indirect human impact on the environment and its inhabiting plants and animals). The Upper Narew watershed is an example of land use that leads to environmental degradation and unfavorable changes, such as the imbalance of water in the soil as a result of dredging the drainage of rivers, or defective melioration. A calculation can be made for $c = 0$ or alternatively, the value of the lower bound can be estimated by another method (e.g. empirically) or from a chart of the normalized variable u by creating a straight line on a normal probability distribution plot.

3.3.1. Determination of statistical characteristics of $Q_{max,f}$ and V_f series

As the estimation of copula-generated probability distribution parameters requires optimal marginal distributions, the first step is to investigate those probabilistic or randomness properties of $Q_{\mathrm{max,f}}$ and V_{f} (e.g.: mean value, median, standard deviation, variance, skewness and kurtosis) that foster selection of theoretical models for a 1D probability distribution. This analysis is very useful in selecting – on the basis of observed values of the random variables $Q_{\mathrm{max,f}}$ and V_{f} – both the accurate probability density functions $f(q_{\mathrm{max,f}})$ and $f(v_{\mathrm{f}})$, and the accurate cumulative distributions.

3.3.2. Models of probabilistic properties of $Q_{max,f}$ and V_f

It was assumed that three possible mathematical models were capable of describing the probabilistic properties of the random variables $Q_{\mathrm{max,f}}$ and V_{f}, namely: (i) log-normal (LN) (Johnson et al., 1994), gamma (GA) (Aksoy, 2000), Weibull (WE) (Sagias and Karagiannidis, 2005), (ii) Inverse Gaussian (IGa) (Barndorff-Nielsen, 2007), and (iii) Generalized Exponential (GE) (Gupta and Kundu, 1999). In the cases of LN, WE and GE, distribution quantiles can be easily calculated; however, the complex forms of the GA and IGa distribution density functions do not allow one to obtain an analytical form of quantiles. In such cases, theoretical quantiles are obtained by a numerical Newton method that ensures good convergence and short computation time.

3.3.3. Estimation of parameters of marginal probability distribution functions

For LN, GA, WE, IGa and GE distributions, the so-called lower (left-side) bound, d_i, of the random variable is not subject to estimation with regards to the three parameter density functions. Two remaining parameters (scale and shape) for a preset value of lower bound were estimated by the MLM method. Estimators of parameters of GA, WE and GE distributions were determined from a set of equations by minimizing the suitable formula using the numerical methods of Brent and Newton (Chapra and Canale, 2006). It was assumed that for the LN, GA, WE, IGa and GE distributions, the lower bound d_i satisfies the condition $0 \le d_i < \min_{1\le j\le N}\left(x_j\right)$, where N is the size of the random variable sample, and d_i is allowed to adopt values from 0 to the minimum value of the random variable of the sample. The validity of the d_i value for each distribution was assessed using the Akaike Information Criterion (AIC) as follows (Akaike, 1974):

$$AIC = 2K - 2\ell\left(\hat{\theta}\right) \tag{21}$$

where,

$\ell\left(\hat{\theta}\right)$ is the logarithm of the likelihood function for the estimated vector of parameters $\hat{\theta}$, and

K is the number of parameters of the density function.

According to AIC, the best model with lower bound d_i is the one for which the AIC value is lowest.

3.4. Selection of the best fitted marginal probability distribution for random variables $Q_{max,f}$ and V_f

The selection of the best fitted probability distribution focuses on the tails of tested distributions. Tail regions are very important from the point of view of the occurrence of extreme values of a random variable (i.e. values with very small probabilities). Therefore, the assessment of the goodness-of-fit marginal distribution employed a number of statistical tests: Kolmogorov–Smirnov (K–S) (Genest et al., 2006; Genest et al., 2009), Anderson–Darling (A–D), Liao–Shimokawa (L–S) (Liao and Shimokawa, 1999) and Kuiper (K) (Koziol, 2008).

The K–S test can be used in verifying large deviations of a theoretical distribution from an empirical one. The A–D test is sensitive to deviations in the tail region (Abidin et al., 2012), while the L–S test is the best test to verify Gumbel and 2-parameter Weibull distributions (Liao and Shimokawa, 1999). A K test was employed in order to verify the goodness-of-fit of marginal distributions in the region of the median as well as in the lower and upper tails of a distribution. Values of calculated statistics for every tested distribution were compared with critical values computed by a bootstrap method (Amal, 2006). For further analysis, these distributions were selected when the AIC criterion was minimal, rendering the null hypothesis regarding the goodness-of-fit between the theoretical and empirical distributions unable to be rejected ($\alpha = 0.05$).

3.5. Estimation of parameters of 2D normal distribution and generating realizations of random variables ($q_{max,f}$, v_f)

Parameter estimation for the 2D normal distribution followed the MLM method (Kotz et al., 2000) and yielded both an estimator of the expected value matrix $\hat{\boldsymbol{\mu}} = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ as well as an unbiased estimator of the covariance matrix $\hat{\Sigma} = \begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}$.

The 2-dimensional random variable ($Q_{max,f}$, V_f) realization was generated in three steps: (i) through a Cholesky decomposition of the 2×2 dimension $\mathbf{\Sigma}$ matrix, where $\mathbf{C}$ was the upper triangle 2×2 dimension matrix, and $\mathbf{C}^T$ was the transposition matrix; a suitable matrix $\mathbf{C}$ was built to meet the condition $\mathbf{CC}^T = \mathbf{\Sigma}$, (ii) the realization of the 2D random variable of the standard normal distribution $\boldsymbol{z} = (z_1, z_2)^T$ with $k = 1000$ elements was generated using a Box–Muller transformation (Box and Muller, 1958), and (iii) the random variable $\boldsymbol{Z}$ was transformed according to the formula $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{A}\,\boldsymbol{z}$.

3.6. Estimation of parameters of a copula functionand generating the random variable $(q_{max,f}, v_f) = (F_1(q_{max,f}), F_2(v_f))$

The most difficult problem in implementing copula theory is the estimation of unknown parameters of marginal probability distributions as well as the copula itself. Parameter estimations occurred in two stages: (i) estimation of parameters of the marginal probability distributions by the MLM method, and (ii) estimation of the correlation coefficient ρ by Bayesian simulation for the elliptical Gaussian copula and by the IFM method for the Archimedean 1-parameter Gumbel–Hougaard copula.

3.6.1. Estimation of elliptical Gaussian copula

In the case of the 2D elliptical Gaussian copula, estimation of the single parameter correlation coefficient ρ was achieved through a Monte Carlo simulation method (Danaher and Smith, 2009), based on Markov Chain Monte Carlo (MCMC). The algorithm was carried out in five steps. In the first step, random variables $Q_{max,f}$ and V_f were transformed twice according to equations:

$$x^*_{1,i} = \Phi^{-1}[F_1(q_{\max,i})] \text{for} i = 1, ..., N \tag{22}$$

$$x^*_{2,i} = \Phi^{-1}[F_2(v_{f,i})] \text{for} i = 1, ..., N \tag{23}$$

where,

F_1 and F_2 are cumulative distribution functions of margins, and

Φ^{-1} is the reverse function to the cumulative distribution of the normal distribution.

In this manner, the set of realizations of random variables$x^* = \left\{x^*_{1,i}, x^*_{2,i}\right\}$ was obtained for $i = 1,\ldots,N$, generated by the means of an elliptical Gaussian copula and transformed to an $R^2 = R^+ \times R^+$ space. The set of observed realizations of random variables Q_{max} and V_f was denoted as $x = \left\{q_{\max,i}, v_{f,i}\right\}$ for $i = 1,\ldots,N$ where $N = 44$.

In the second step, the correlation coefficient r, conditioned by sets x^* and x, was estimated by the Metropolis-Hastings method (Chib and Greenberg, 1995). This involves the generation of a new variable value r^{new} by means of a normal distribution, where $r^{new} \sim N(r^{old}, 0.01)$ and r^{old} are the values of the correlation coefficient obtained in the previous iteration of the Marcov chain on the basis of criterion α:

$$\alpha = \min(1, \frac{f(r^{\text{new}}|r, x^*, x)}{f(r^{\text{old}}|r, x^*, x)}) \tag{24}$$

This criterion is connected to a conditional *a posteriori* distribution (Danaher and Smith, 2009). The value of the random variable $u \sim U(0,1)$ from the uniform distribution was compared with criterion α. If $u \le \alpha$ then $r = r^{new}$; otherwise $r = r^{old}$.

If R is a 2×2 dimensional upper triangle matrix with correlation coefficients (r) and estimated by the Metropolis-Hastings method (as its elements are beyond the main diagonal that consists of elements equal to 1), then the Cholesky decomposition of the $2{\times}2$ dimension $\mathbf{\Sigma}$ matrix follows the equation $\Sigma = \left(R^T R\right)^{-1}$, where R^T is the transposition matrix. The 2×2 dimension $\Gamma = \text{diag}(\Sigma)^{-1/2}\, \Sigma\, \text{diag}(\Sigma)^{-1/2}$ matrix is then computed, where diag(Σ) becomes the diagonal matrix comprised of the leading diagonal of $\mathbf{\Sigma}$.

In the third step, the value of the correlation coefficient ρ serves as the input to a randomizing realization of variables procedure that yields the normal distribution $\boldsymbol{Z} \sim N(0, \rho)$. In the fourth step, the realization of the 2D random variable is generated by means of a Gaussian copula as $\boldsymbol{V} = (\Phi(Z_1), \Phi(Z_2))$. The fifth and final step consists of transforming the random variable $\boldsymbol{V}$ to the $R^2 = R^+ \times R^+$ space i.e. $\boldsymbol{X} = (F_1^{-1}(V_1), F_2^{-1}(V_2))$.

During the study, $k = 10{,}000$ iterations of the algorithm were run. A set of random variables $\{(P^{(1)}, \boldsymbol{X}^{*(1)}, \boldsymbol{V}^{(1)}, \boldsymbol{X}^{(1)}), \ldots, (P^{(k)}, \boldsymbol{X}^{*(k)}, \boldsymbol{V}^{(k)}, \boldsymbol{X}^{(k)})\}$ was obtained. An estimator $\bar{\rho}$, of the correlation coefficient ρ, computed by the algorithm of the Metropolis-Hastings method, was calculated as (Chib and Greenberg, 1995):

$$\bar{\rho} = \frac{1}{k}\sum_{i=1}^{k} \rho^{(i)} \tag{25}$$

The first elements of the chain were skipped in the estimation (Eq. (25)) because their probability distribution was biased by the influence of the starting point $\rho^{(0)}$ (Barker and Kelsey, 2012).

3.6.2. Estimation of Archimedean 1-parameter Gumbel–Hougaard copula

The estimation of the Archimedean 1-parameter Gumbel–Hougaard copula was carried out in two steps using the IFM method (McLeish and Small, 1988; Cherubini et al., 2004). In the first step, the parameters θ_j of the marginal probability distribution were estimated as:

$$\hat{\theta}_j = argmax_{\theta_j} \sum_{t=1}^{N} \log f_j(x_{jt}; \theta_j) \quad \text{or} \quad j = 1.2 \tag{26}$$

where:

N is the size of random sample.

On the basis of the estimated parameters $\hat{\theta}_j$, in the second step, the parameters of the copula Φ were estimated as follows:

$$\hat{\Phi} = argmax_{\Phi} \sum_{t=1}^{N} \log c(F_1(x_{1t}, \hat{\theta}_1), F_2(x_{2t}, \hat{\theta}_2)) \tag{27}$$

The realization of the 2D random variable ($Q_{\max,f}$, V_f) was generated using a Gumbel–Hougaard copula generator (Cherubini et al., 2004; Matúš, 2009):

$$\phi(u) = (-\log u_1)^{\theta} \tag{28}$$

$$\phi^{-1}(t) = \exp(-t^{1/\theta}) \tag{29}$$

$$\phi^{-1(1)}(t) = -\exp\left(-t^{1/\theta}\right)\frac{1}{\delta}t^{\frac{1}{\theta}-1} \tag{30}$$

drawing values of variable v_1 and v_2 from the range [0,1] with a uniform distribution, and $u_1 = v_1$,

$$c_1 = \phi(u_1) = (-\log u_1)^{\delta} \tag{31}$$

$$c_2 = \phi(u_1) + \phi(u_2) = (-\log u_1)^{\delta} + (-\log u_2)^{\delta} \tag{32}$$

$$v_2 = c_2(u_2|v_1)\,\text{i.e.,}\ v_2 = \frac{\phi^{-1(1)}(c_2)}{\phi^{-1(1)}(c_2)} \tag{33}$$

thus,

$$v_2 = \frac{-\frac{1}{\theta}\exp\left(-\left[(-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right]^{\frac{1}{\theta}}\right)\left[(-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right]^{\frac{1}{\theta}-1}}{-\frac{1}{\theta}\exp\left(-\left[(-\log u_1)^{\theta}\right]^{\frac{1}{\theta}}\right)\left[(-\log u_1)^{\theta}\right]^{\frac{1}{\theta}-1}} \tag{34}$$

3.7. Measures of goodness-of-fit for 2D probability distribution

When estimating parameters of a 2D probability distribution by the copula method, more than one joint function is generally taken into account, so that the problem of selecting an appropriate copula type is crucial, requiring one or more classification criteria. Five measures were used to verify the goodness-of-fit of the theoretical distribution with the empirical distribution (Table 1): Kolmogorov–Smirnov (K–S) (Genest et al., 2006; Genest et al., 2009), Anderson–Darling (A–D) (D'Agostino and Stephens, 1986), Integrated Anderson–Darling (IA–D) (Gwinn, 1993; Kotz et al., 2001), Kuiper distance (K) (Stephens, 1969; Koziol, 2008) and Euclidean distance (L^2) (Gower, 1985). The above-mentioned measures were adapted for the 2-dimensional case. Copula measures of goodness-of-fit were obtained by computing the distance between the empirical copula C_{emp} and the parametric copula C_{teo} fitted to the data (Mendez et al., 2007).For the K–S test, the measure of goodness-of-fit is the maximal distance between the theoretical and empirical cumulative distribution for various regions of the probability distribution. Having minimal maximal-distance points between theoretical and empirical data distributions indicates which theoretical distribution function best fits the empirical data. The K–S test draws attention to deviations in the central part of the distribution. Consequently, the A–D test was applied to measure deviations in the tail parts of the distributions. As the IA–D test is less sensitive to large deviations, its statistic reduces the influence of the occurrence of outliers in the sample. The statistics A–D and IA–D emphasize deviations in the tails (i.e. the corners of the unit square) by applying a weight function (Mendez et al., 2007) $w = 1/\sqrt{C_{\text{teo}}(u_i, v_j)\left(1 - C_{\text{teo}}(u_i, v_j)\right)}$ to the K–S formula, where: $C_{\text{teo}}(u_i, v_j)$ is a parametric copula and $i, j = 1, \ldots, N$ is the size of the random sample. The K test checks deviations over the entire range of the probability distribution, but with special attention to the median region and the tails. An alternative solution to these tests is the Euclidean distance L^2, which measures the distance between two points in geometric space. In this case, there is no influence of outliers in the sample on the ultimate result of the verification.

Table 1
Goodness-of-fit tests for 2D probability distributions of random variable ($Q_{max,f}$, V_f).

Kołmogorov-Smirnov (K–S)	$D_{K-S} = \max_{1\le i,j\le N} \left\| C_{emp}\left(u_i, v_j\right) - C_{teo}\left(u_i, v_j\right) \right\|$, where: N—size of random sample, $C_{emp}(u,v)$—empirical copula $C_{teo}(u,v)$—theoretical copula $\hat{C}_{emp}(u,v) = \frac{\#\left\{(x_k, y_k): F_X(x_k)\le u, F_Y(y_k)\le v\right\}}{N}$, where: $(x_k, y_k)_{k=1}^{N}$ random sample of size N # number of elements in set, $F_X(x_k)$ F and $_y(y_k)$ marginal empirical probablistic distribution.
Anderson-Darling (A–D)	$D_{A-D} = \max_{1\le i,j\le N} \frac{\left\| C_{emp}\left(u_i, v_j\right) - C_{teo}\left(u_i, v_j\right) \right\|}{\sqrt{C_{teo}\left(u_i, v_j\right)\left(1 - C_{teo}\left(u_i, v_j\right)\right)}}$
Integrated Anderson-Darling (IA–D)	$D_{IA-D} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\left(C_{emp}\left(u_i, v_j\right) - C_{teo}\left(u_i, v_j\right)\right)^2}{C_{teo}\left(u_i, v_j\right)\left(1 - C_{teo}\left(u_i, v_j\right)\right)}$
Kuiper (K)	$D_K = \max_{i,j}\left(C_{emp}\left(u_i, v_j\right) - C_{teo}\left(u_i, v_j\right)\right) + \max_{i,j}\left(C_{teo}\left(u_i, v_j\right) - C_{emp}\left(u_i, v_j\right)\right)$
Euclidean distance (L^2)	$D_{L^2} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(C_{emp}\left(u_i, v_j\right) - C_{teo}\left(u_i, v_j\right)\right)^2}$

Table 2
Upper and lower indices of probability distribution tails.

Copula			Symbol	τ^l	τ^u	Remarks
1-parameter copula	Gaussian	Elliptical	B0	0	0	$C^-_{\rho=-1}, C^\perp_{\rho=0}, C^+_{\rho=1}$, where: $C^- = \min(v+z)$—minimum copula $C^+ = \max(v+z-1.0)$—maximum copula $C^\perp = v \cdot z$—multiplicative copula
	Gumbel–Hougaard	Extreme Values EV	B6	0	$2-2^{\frac{1}{\delta}}$	$C^\perp_{\delta=1}, C^+_{\delta\to+\infty}$ Upper tail is heavy, Lower tail is light (normal)

In the methodology for selecting a mathematical model, emphasis was put on the tail portions of the distributions because it is in these regions that the maximum values of Q_{max} and V_f occur. The method of parameter estimation for a 2D probability distribution by means of a copula enables the determination of the degree of 'heaviness' of heavy-tails as well as the relationship between tails of marginal cumulative distributions. The relationship between extreme values was described by the lower τ^l and upper τ^u indices of tails of the probability distribution (Table 2).

4. Results and discussion

4.1. Identification of snowmelt floods for the Upper Narew River at Wizna, Poland

44 snowmelt floods were identified among the series of discharge measurements recorded at the Wizna station between 1966 and 2012. For all seasonal flood events not classified as snowmelt floods (i.e., 1972, 1975, 1981, 1984, 2001 and 2008), the snow cover was not very deep and developed over a short period of time, mainly in the eastern portion of the Biebrza River watershed and in the upper portion of the Narew River watershed. However, in 1986, 2005 and 2011, two snowmelt floods occurred. For the 44 accepted snowmelt flooding events, the 2D random variable ($Q_{max,f}$, V_f) sample was constructed taking into consideration peak $Q_{max,f}$ [$m^3\ s^{-1}$] and volume V_f [$10^6\ m^3$] of each flood (Fig. 2).

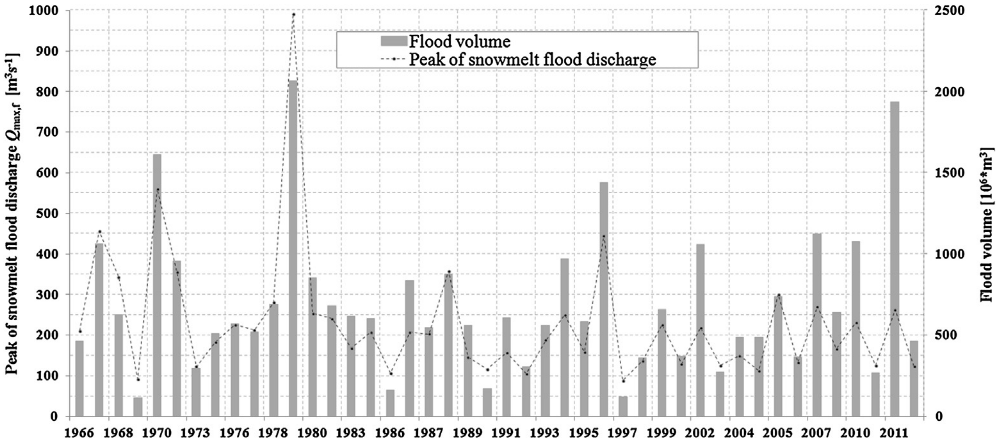

Fig. 2. Maximum flow and volume of snowmelt floods for the Upper Narew River at Wizna, Poland (cross-section from the period 1966–2012).

4.2. Homogeneity analysis of $Q_{max,f}$ and V_f

The homogeneity analysis was carried out in accordance with the procedure presented in Section 3.2. While the Grubbs–Beck test (Grubbs and Beck, 1972; Bulletin 17B, 1982) used in homogeneity analysis recognized one outlier (e.g. a high-magnitude flood) in the tested sample of $Q_{\rm max,f}$, genetic analysis of this flood showed it to conform to the natural behavior of flood events and it was therefore not rejected from the sample. No outliers were detected in the $V_{\rm f}$ sample. As the data also passed statistical tests for independence (Anderson, 1941; Bulletin 17B, 1982; Pilon et al., 1985) and stationarity (Dahmen and Hall, 1990; Sneyers, 1990), it was concluded that the homogeneity analysis procedures indicated no reason to reject the null hypothesis regarding the homogeneity of the tested data.

4.3. Statistical analysis of $Q_{max,f}$ and V_f

4.3.1. Normalization and verification process

Logarithmic transformations with various lower limits c were employed for normalization. The Kolgomorov–Smirnov test was used to verify the normality marginal distributions of normalized and original values of $Q_{\rm max,f}$ and $V_{\rm f}$. The critical values of the test were computed using a bootstrap method (Amal, 2006). While the Kolgomorov–Smirnov test rejected the hypotheses of normality for the original variables $Q_{\rm max,f}$ and $V_{\rm f}$ ($\alpha = 0.05$), the hypothesis of normality for normalized variables with various limits c was not rejected, with the lowest test statistics obtained for the normalized variables $Q_{\rm max,f}$ [$c = 41.94\,{\rm m^3 s^{-1}}$]; (i.e. $\ln(Q_{\rm max,f} - 41.94)$), and $V_{\rm f}$ [$c = 0\,{\rm m^3}$]; (i.e. $\ln(V_{\rm f})$).

The normal probability plots for $\ln(Q_{\rm max,f} - 41.94)$ and $\ln(V_{\rm f})$ confirmed the results of the Kolmogorov test. Assessing the goodness-of-fit hypothesis for tested variables' distributions with normal distributions requires co-linearity of both theoretical and empirical distributions. If the observed values of random variables (abscissa on plots) yield a normal distribution, then all points should be close to a straight line (Figs. 3 and 4). Parameters of the normal probability distribution for the 2D random variable $\left(\ln(Q_{\rm max,f} - 41.94), \ln(V_{\rm f})\right)$ were estimated by the ML method.

4.3.2. Characteristics of the probability distribution for $Q_{max,f}$ and V_f

To compare a 2D normal probability distribution with a 2D copula-based probability distribution, it was necessary to determine the best fitted marginal distributions for variables $Q_{\rm max,f}$ and $V_{\rm f}$. For this, the measures of asymmetry, peakedness, dispersion and position were calculated for variables $Q_{\rm max,f}$ and $V_{\rm f}$ as well as their normalized counterparts.

A large positive skewness of $Q_{\rm max,f}$ indicated that the tail on the right side of the probability density function was longer or 'fatter' than the one on the left side. In the case of variables $\ln(Q_{\rm max,f} - 41.94)$ and$\ln(V_{\rm f})$, skewness ranged between ± 1.5, indicating that their distributions were symmetrical and close to that of a normal distribution.

The positive kurtosis value for the original variable $Q_{\rm max,f}$ indicated a heavier tail of its distribution than that of a normal distribution. However, small positive values of kurtosis (± 3.0) for $\ln(Q_{\rm max,f}{-}41.94)$ and $\ln(V_{\rm f})$ indicated that their distributions were close to a normal distribution, but remained slightly more slender. In the case of $V_{\rm f}$, the skewness and kurtosis were 1.352 and 1.774 respectively, indicating a close-to-normal distribution; however, a comparison of mean and median contradicted this result.

4.4. Selection of the best fitted function of the marginal distribution for $Q_{max,f}$ and V_f

Five probability distributions—log-normal (LN), gamma (GA), Weibull (WE), Inverse Gaussian (IGa) and Generalized Extreme Value (GE)—were employed to select the best fitted marginal PDF. The best fitted distribution was selected when

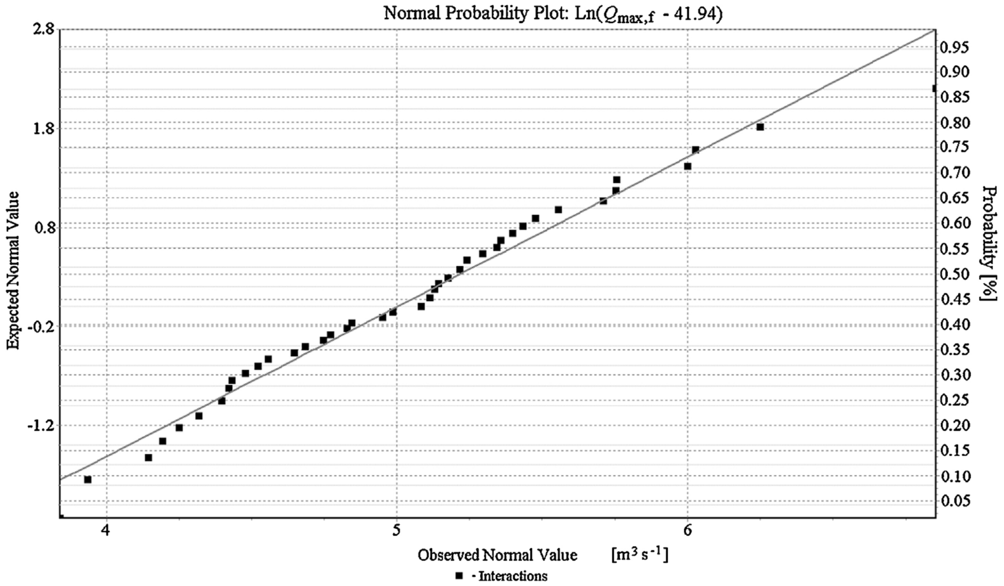

Fig. 3. Normal Probability Plot for $\ln(Q_{max,f}-41.94)$.

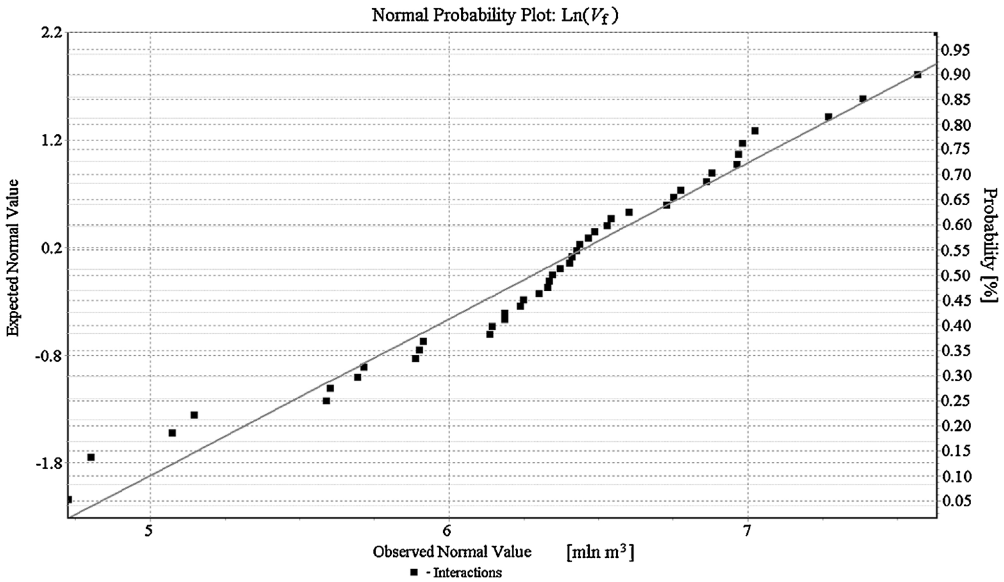

Fig. 4. Normal Probability Plot for $\ln(V_f)$.

the AIC criterion was minimal and the A–D, K–S, L–S and K statistical tests ($\alpha = 0.05$) did not reject the goodness-of-fit hypothesis. In close relation to the size of the random sample being based upon the parameters of the estimated PDF, the significance level was also based on the authors' experiences in testing theoretical and empirical distributions. Generally, the larger the sample is, the lower the value of the significance level that can be adopted. Given a sample size of 44 for the tested variables, the acceptable significance level would be $\alpha = 0.01$; however, to avoid committing a type I error (i.e. rejection of the tested null hypothesis H_0 when it is true), $\alpha = 0.05$ was adopted. Nevertheless, critical values of tests for $\alpha = 0.01$ are also presented in Tables 3 and 4.

For $Q_{max,f}$, the tests rejected the GE distribution and for V_f, the tests rejected the LN (K–S test) and the IGa (A–D, K–S, L&S and K tests)($\alpha = 0.05$). Based on the AIC criterion (Tables 3 and 4), the LN ($AIC_{LN} = 528.24$) and WE ($AIC_{WE} = 647.34319$) distributions were selected as the best fitted for $Q_{max,f}$, and V_f respectively. Since values of AIC accepted by the test distributions differed slightly, it was decided that a graphical comparison of the theoretical and empirical distributions would be conducted as an additional assessment of goodness-of-fit.

Table. 3
Results of goodness-of-fit tests for the $Q_{max,f}$ variable for the Akaike Information Criterion (AIC). Bold features indicate the best fitted theoretical probability distribution, while underlined features show distributions for which tests rejected the null hypothesis at $\alpha = 0.01$ and/or $\alpha = 0.05$.

	A-D	K-S	L&S	K	AIC
LN	0.21768	0.09774	0.58297	0.14509	**528.23671**
$\alpha_{cr.} = 0{,}05$	0.63806	0.12246	0.90300	0.21111	
$\alpha_{cr.} = 0{,}01$	0.86145	0.14291	0.94938	0.24126	
GA	0.31316	0.07785	0.70813	0.14020	528.72509
$\alpha_{cr.} = 0{,}05$	0.67617	0.11506	0.94938	0.21179	
$\alpha_{cr.} = 0{,}01$	1.00582	0.13930	1.15782	0.24261	
WE	0.41509	0.07798	0.80344	0.15263	528.56699
$\alpha_{cr.} = 0{,}05$	0.81890	0.13268	1.04046	0.22419	
$\alpha_{cr.} = 0{,}01$	1.09537	0.15665	1.23683	0.25477	
IGa	0.26124	0.10214	0.64004	0.15342	528.70531
$\alpha_{cr.} = 0{,}05$	0.65858	0.12570	0.91480	0.21090	
$\alpha_{cr.} = 0{,}01$	0.94334	0.14985	1.05684	0.24442	
GE	16.18132	0.34217	5.32600	0.36416	567.35627
$\alpha_{cr.} = 0{,}05$	0.80389	0.13701	1.01582	0.22201	
$\alpha_{cr.} = 0{,}01$	1.20193	0.16646	1.24568	0.25656	

Table 4
Results of goodness-of-fit tests for V_f variable for the Akaike Information Criterion (AIC). Bold features show the best fitted theoretical probability distribution, while underlined features indicate distributions for which tests rejected the null hypothesis at $\alpha = 0.01$ and/or $\alpha = 0.05$.

	A–D	K–S	L&S	K	AIC
LN	0.53710	0.12790	0.80442	0.18860	649.25300
$\alpha_{cr.} = 0{,}05$	0.72067	0.12720	0.93932	0.21393	
$\alpha_{cr.} = 0{,}01$	0.93204	0.14998	1.08232	0.24734	
GA	0.42637	0.10842	0.76940	0.19467	647.58226
$\alpha_{cr.} = 0{,}05$	0.70727	0.11822	0.98340	0.21361	
$\alpha_{cr.} = 0{,}01$	0.99073	0.13809	1.27119	0.24831	
WE	0.49901	0.11137	0.84450	0.20571	**647.34319**
$\alpha_{cr.} = 0{,}05$	0.82052	0.13089	1.03031	0.22181	
$\alpha_{cr.} = 0{,}01$	1.08853	0.15512	1.19596	0.24998	
IGa	0.79429	0.15360	0.94873	0.21135	650.74591
$\alpha_{cr.} = 0{,}05$	0.75147	0.13437	0.96640	0.21563	
$\alpha_{cr.} = 0{,}01$	0.97969	0.16031	1.13220	0.25285	
GE	0.39657	0.09879	0.74553	0.18583	647.73770
$\alpha_{cr.} = 0{,}05$	0.93225	0.14807	1.04318	0.22376	
$\alpha_{cr.} = 0{,}01$	1.48164	0.18732	1.31003	0.25964	

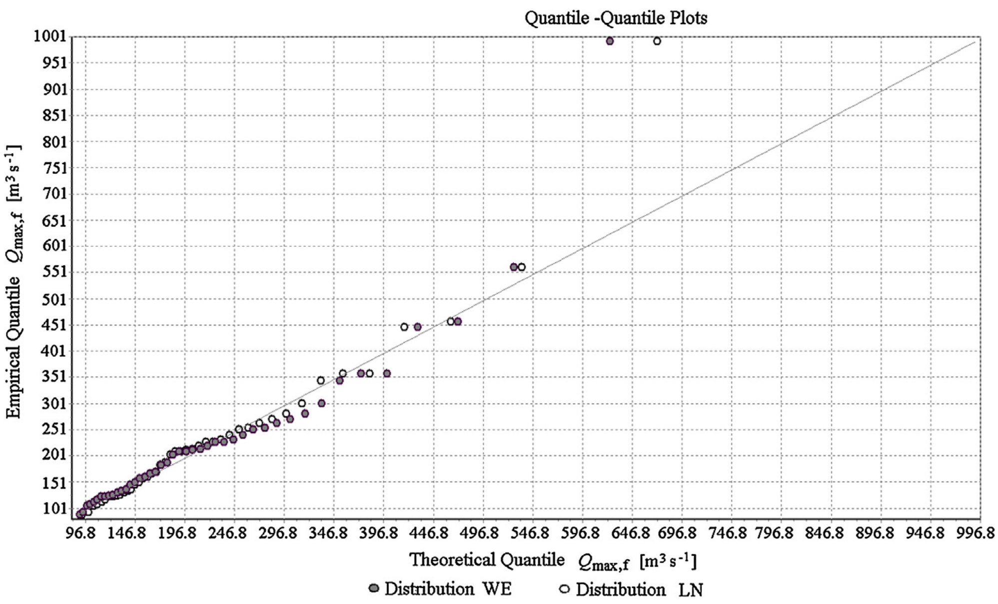

Fig. 5. Quantile theoretical—Quantile empirical (Q–Q) Plot for $Q_{max,f}$.

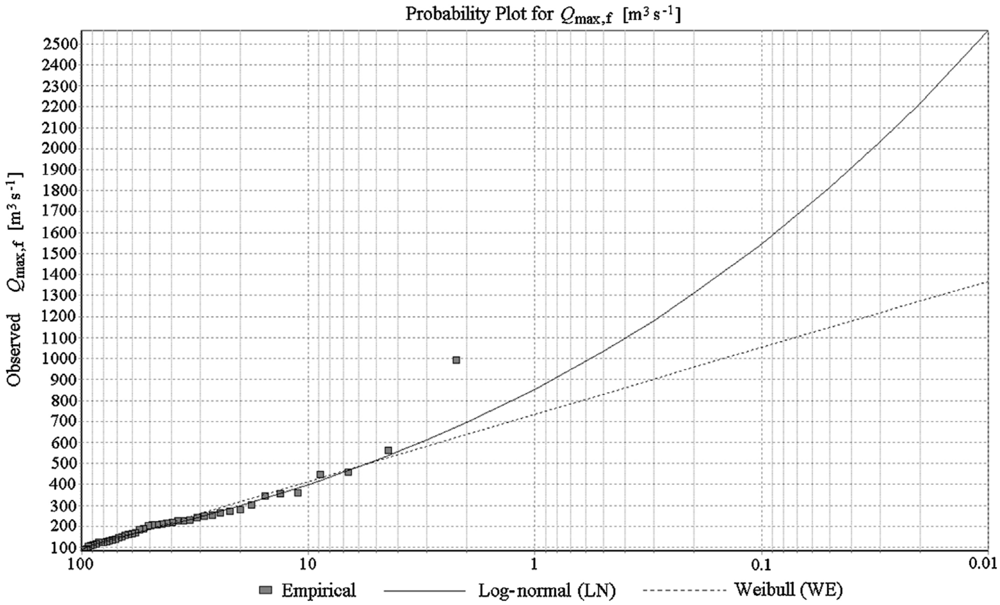

Fig. 6. Probability Plot for $Q_{max,f}$.

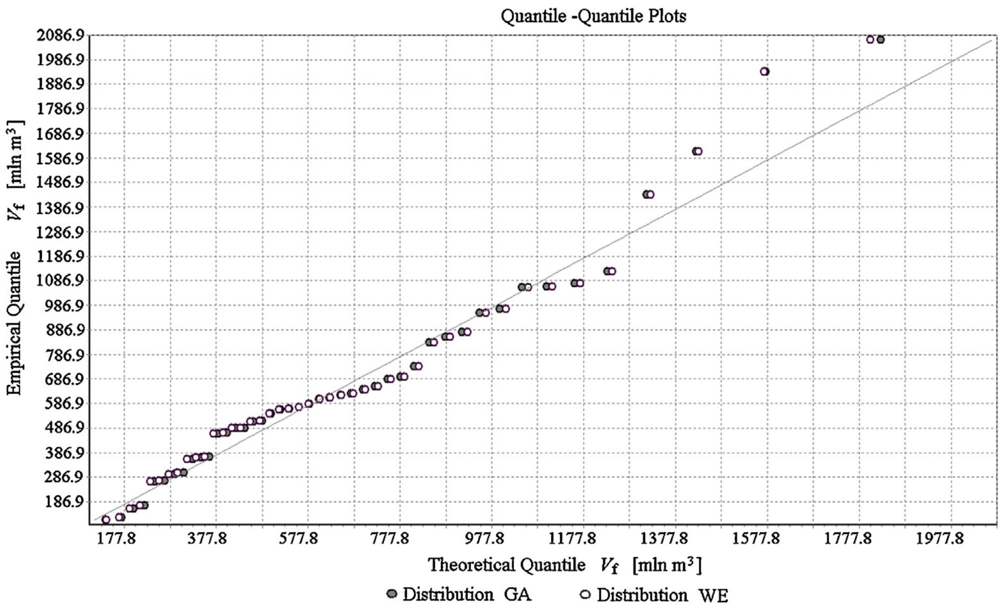

Fig. 7. Quantile theoretical—Quantile empirical (Q–Q) Plot for V_f.

A *q–q* plot (e.g. empirical quantiles–theoretical quantiles) is presented in Fig. 5, where observed values of $Q_{max,f}$ versus expected values from WE (e.g. first choice Table 3) and LN (e.g. second choice Table 3) distributions were drawn. An ideal fit of the theoretical distribution to observed values would result in a straight line relationship for the plots. Therefore, based on a visual assessment of the plots, both WE and LN distributions were deemed to approximate the empirical distribution well. The WE distribution had a lighter tail (e.g. coloured points deviated more from a straight line than did the blank points of the LN distribution, showing the latter to be a better fit). Moreover, the comparison of WE and LN probability distribution plots for $Q_{max,f}$ (Fig. 6), showed the LN distribution to better approximate observations than the WE distribution, in particular for the upper region where extreme values of snowmelt flood occur (e.g. $Q_{max,f}$ = 992.0 m^3 s^{-1} in 1979).

A similar graphical analysis was conducted for the V_f variable. For this variable, the minimal values of AIC criterion were obtained for WE (e.g. first choice, Table 4) and GA (e.g. second choice, Table 4) distributions. The probability distribution plots for both distributions were very similar (Fig. 7). For the upper portion of distributions, the last element of the GA distribution is slightly closer to a straight line than the last element of the WE distribution. This was confirmed by the smaller value of

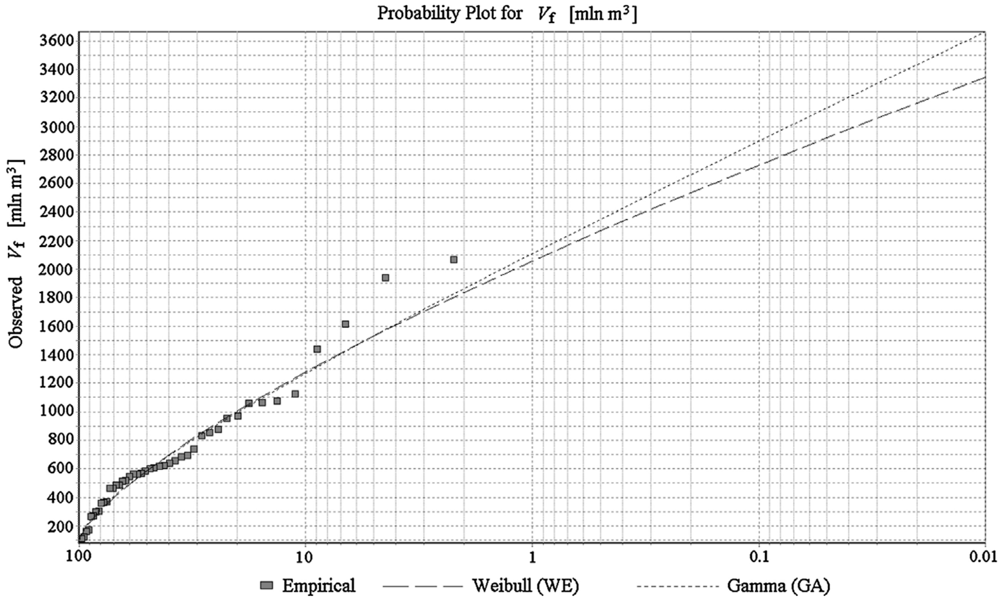

Fig. 8. Probability Plot for V_f.

A–D test statistics for the GA *vs.* WE distributions (A–D$_{GA}$ = 0.426 *vs.* A–D$_{WE}$ = 0.499 (Table 4)). However, for the remaining quantiles, the WE distribution is closer to a straight line. Because the right tail of the WE distribution is slightly lighter than for the GA distribution, the selection of the WE distribution gave a smaller probability of occurrence of extreme V_f values (Fig. 8). Lastly, on the basis of the entire analytical and graphical assessment of goodness-of-fit, the LN distribution and WE distributions were chosen for $Q_{max,f}$ and V_f as the best fitted marginal distributions for 2D probability distribution estimation via the copula method.

4.5. *Comparison of 2D probability distribution functions of normal distribution and Gaussian and Gumbel–Hougaard copulas for the ($Q_{max,f}$, V_f) variables*

Prior to comparison of their 2D normal *vs.* copula-based distributions, the relationship between the original random variables $Q_{max,f}$ and V_f was investigated. A linear regression of $Q_{max,f}$ *vs.* V_f with

$$V_f = 2.28Q_{max,f} + 162.50 \quad (35)$$

indicates a strong linear correlation ρ_p = 0.813 between the two variables, with some regions of nonlinearity in the plot. In the case of a monotonic relationship, the nonparametric Spearman correlation can be used. A Spearman correlation coefficient of ρ_s = 0.900 confirmed the high level of co-dependence between the tested variables, confirming the appropriateness of using a 2D normal probability distribution for the probabilistic description of the ($Q_{max,f}$, V_f) variable.

4.5.1. *Parameters of 2D normal probability distribution, Gaussian and Gumbel–Hougard copula functions and measures of goodness-of-fit*

An important step in the construction of the 2D model for 2D variables ($Q_{max,f}$, V_f) is assessing suitable 2D copula functions in comparison with their observations. In this study, two well-known families of copulas were arbitrarily chosen, featuring a wide range of dependence, and representing elliptical and Archimedean families and a binormal probability distribution function. In Figs. 9–11 , the level curves of the empirical copula were compared to the theoretical bivariate normal distribution, Gaussian and Gumbel–Hougaard copulas, respectively. They were fitted to the available observations, with those of the empirical copula constructed using the same data. These above comparisons were carried out for the following quantiles: 0.25, 0.50 and 0.75. The straight lines obtained and jumps observed in the empirical curves are due to the presence of identical pairs of observed values, as well as the finite, discrete nature of the sample values. Occurrence of such events spoils the estimation of chosen probabilities.

The above can be performed as a comparison of the distribution of ranks (i.e. properly normalized into a domain of unity I) (De Michele et al., 2007; Salvadori et al., 2007). Marginals are not used to construct the empirical copulas and the analyzed domain is the unit square $\boldsymbol{I}^2$. In this way, we can obtain a non-parametric bivariate distribution. However, in this study it was decided to use a procedure that allows us to compare the probability curves of theoretical and empirical distributions in the space filled with observations (black points) and a cloud of 2D variables generated from the tested bivariate distributions (gray points) in Figs. 9–11.

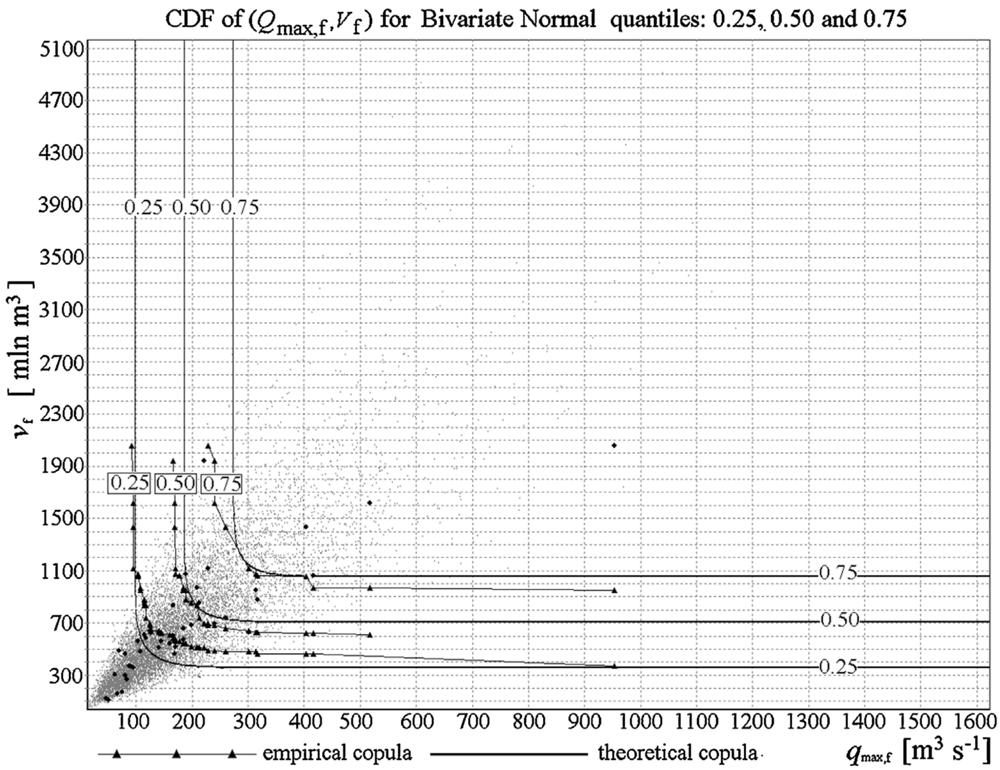

Fig. 9. Comparison between the level curves of the theoretical bivariate normal distribution (solid lines), fitted to the available observation (black circle) and those of the empirical distribution (lines with triangles) constructed using the same data.

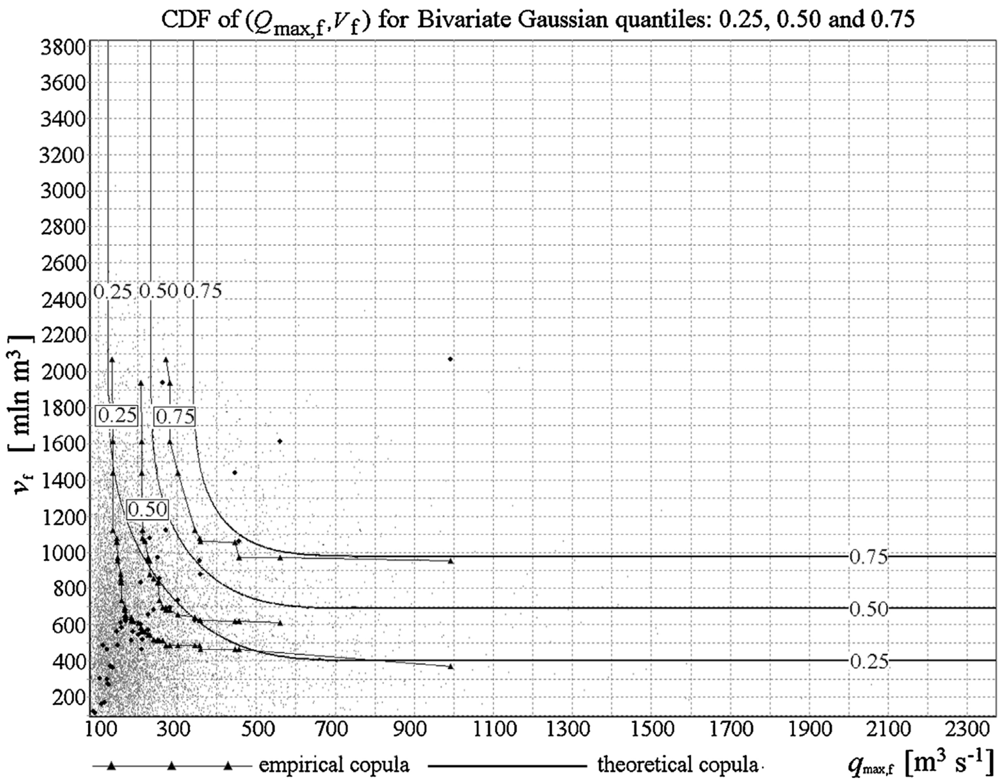

Fig. 10. Comparison between the level curves of the theoretical bivariate Gaussian copula (solid lines), fitted to the available observation (black circle) and those of the empirical copula (lines with triangles) constructed using the same data.

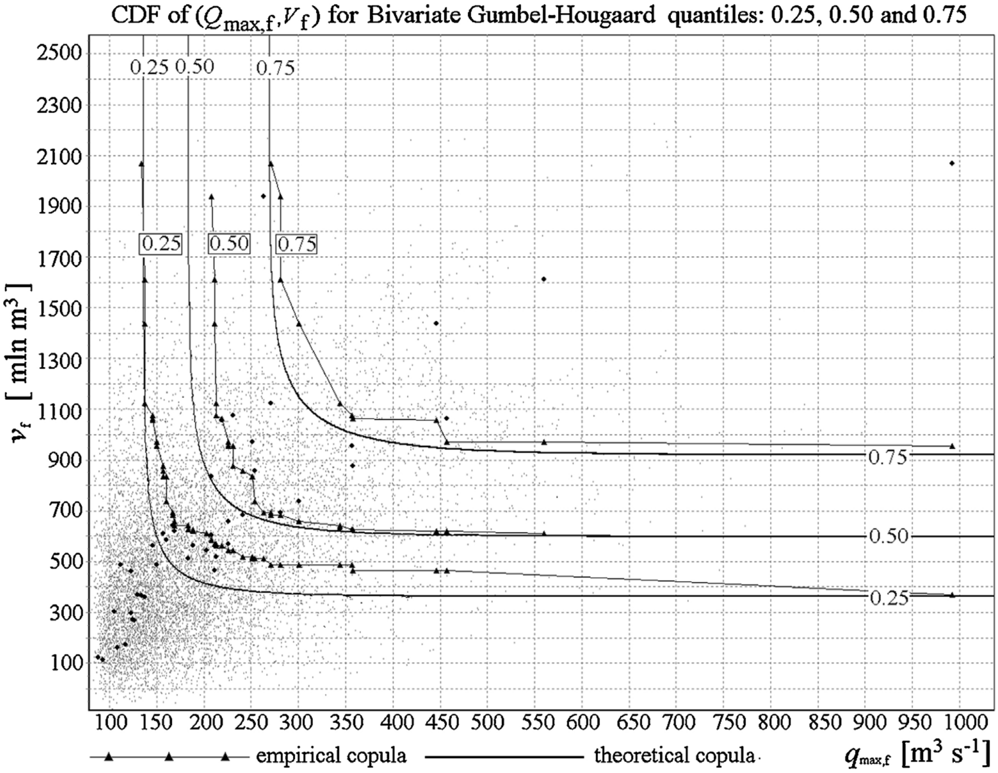

Fig. 11. Comparison between the level curves of the theoretical bivariate Gumbel–Hougaard copula (solid lines), fitted to the available observation (black circle) and those of the empirical copula (lines with triangles) constructed using the same data.

Table 5
Estimators of parameters of 2D probability distributions estimated by MLM, Gaussian and Gumbel–Hougaard copulas, Spearman rank correlation coefficients ρ_s as well as results of K–S, A–D, IA–D, K and L^2 goodness-of-fit tests for the ($Q_{max,f}$, V_f) variables (bold features refer to best fitted distribution according to the specific test).

Estimated parameters of 2D distributions		ρ_s	K–S test	A–D test	IA–D test	K test	L^2 test
2D Normal distribution		0.9809	**0.0361**	0.4803	119.006	0.2727	4.6749
$\bar{\rho}$	0,871						
Gaussian copula		0.7721	0.1368	2.3854	439.292	0.2384	**2.5513**
$\bar{\rho}$	−0,665						
Gumbel–Hougaard copula		0.9361	0.0517	**0.4031**	**63.175**	**0.2296**	3.4492
θ	1.674						
τ_{upper}	0.487						

When the marginal distributions were normal, the Gaussian copula generated evenly distributed points characteristic of a joint standard normal distribution (Fig. 10). In other cases, points were not evenly distributed. Copula theory and Monte Carlo simulations were used to determine the level of heaviness of the tails of the multidimensional cumulative distributions as well as their co-independence. The above analysis showed positive dependence among variables describing the studied snowmelt floods as well as some asymmetries in the dependence structure. The above random property is much better described by the Gumbel–Hougaard copula (Fig. 11) than by the 2D normal distribution (Fig. 9) or Gaussian copula (Fig. 10). The Gaussian copula is obviously symmetric and hence the lower and upper tail dependence coefficients are the same, i.e. $\tau^l = \tau^u = 0$, which is a drawback in the application of this copula's function to describe extreme events such as snowmelt floods.

The symmetric distribution of correlation on both sides of the median means that the same correlation will concern the occurrence of extreme high and low values of each marginal variable. This represents a significant impediment to using simulation floods. Such situations should be taken into account in the descriptions of snowmelt floods. An important property of the Gaussian copula is the disappearing correlation at the ends of the density distribution of these functions (Eq. (18)). This means that during the simulation, the extreme values are independent of each other. In the case of modelling extreme flood events, such an assumption is incorrect and despite the simplicity the Gaussian copula is not the best tool for simulating correlation of rare events such as $Q_{max,f}$ and V_f.

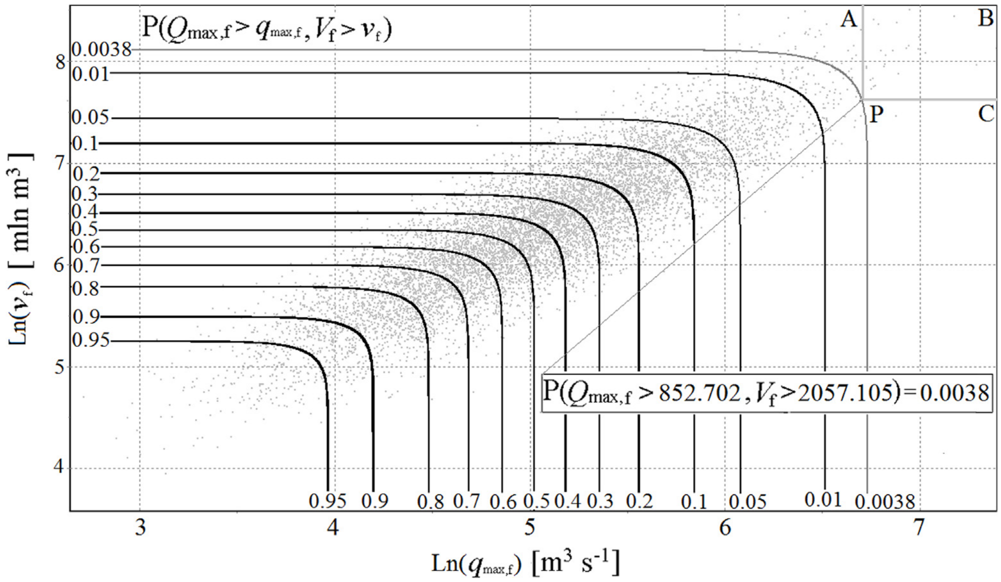

Fig. 12. Exceedance probability function $P(Q_{max,f} > q_{max,f}, V_f > v_f)$ for 2D normal distribution of the $(Q_{max,f}, V_f)$ variables.

Table 5 presents the parameters of the 2D normal distribution estimated by the MLM method, and the 2D probability distributions estimated by the Gaussian and Gumbel–Hougaard copulas, as well as the Spearman rank correlation coefficient ρ_s and five measures of goodness-of-fit. The coefficient ρ_s is equal to 0 for independent variables, and equals 1 or −1, respectively, for ideal increasing or decreasing relationships. A high Spearman rank correlation coefficient was obtained for the normal distribution ($\rho_s = 0.9809$), with slightly lower values for the Gumbel–Hougaard copula ($\rho_s = 0.9361$), and still lower values for the Gaussian copula ($\rho_s = 0.7721$). It is worth noting that the $(Q_{max,f}, V_f)$ variable generated by the Gumbel–Hougaard copula preserved a similar relationship between $Q_{max,f}$ and V_f variables to that of observed values ($\rho_s = 0.900$). The parameter θ of the Gumbel–Hougaard copula is allowed to differ in range [1,+∞). When $\theta = 1$, no relationship exists between the tested variables, whereas for the obtained value of $\theta = 1.674$, a stronger relationship is indicated. The upper index of the upper tail of the Gumbel–Hougaard copula ($\tau^u = 0.487$) indicated that the upper tail was heavier than for the normal distribution (i.e. $\tau^u = 0$).

The selection of the theoretical 2D random variables $(Q_{max,f}, V_f)$ should be carried out with special attention to the tails of the probability distributions given the importance of the occurrence of extreme events. Consequently, since measures of goodness-of-fit were adopted that were sensitive to deviations between theoretical and empirical cumulative distributions, particular attention centered on the outcome of the Anderson–Darling (A–D) and Integrated Anderson–Darling (IA–D) tests. The lowest values of these statistics (e.g. $A–D_{G–H} = 0.4031$ and $IA–D_{G–H} = 63.175$) were obtained for the Gumbel–Hougaard copula with 3-parameter marginal distributions LN and WE, respectively. As a second-best choice, these tests pointed to the 2D normal distributions with parameters estimated by MLM (e.g. $A–D_{Bi–N} = 0.4803$ and $IA–D_{Bi–N} = 110.006$). In the case of the K–S test, the lowest values were also obtained for normal distributions (e.g. $K–S_{Bi–N} = 0.0361$, $K–S_{G–H} = 0.0517$ and $K–S_{GAUSS} = 0.1368$). In contrast to the A–D and IA–D tests, the K–S test is more sensitive to deviations in the central part of the distributions. The Kuiper (K) test, sensitive to deviations in tail regions as well as those close to the median, pointed to the Gumbel–Hougaard copula (e.g. $K_{G–H} = 0.2296$, $K_{GAUSS} = 0.2384$ and $K_{Bi–N} = 0.2727$), whereas the Euclidean distance L^2 that measures the distance between two points (excluding values of outliers) pointed to a PDF built upon a Gaussian copula as the best fitted distribution (e.g. $L^2_{GAUSS} = 2.5513$, $L^2_{G–H} = 3.4492$ and $L^2_{Bi–N} = 4.6749$). The greatest maximum deviations between theoretical and empirical distributions were obtained for a probability distribution with parameters estimated through the use of an elliptical Gaussian copula (e.g. $K–S_{GAUSS} = 0.1368$, $A–D_{GAUSS} = 2.3854$, $IA–D_{GAUSS} = 439.292$). On the basis of the analysis of the nonparametric Spearman rank correlation coefficient ρ_s, the outcomes of tests and visual assessment of graphs, it was decided that 2D normal distribution and 2D probability distribution built using an Archimedean 1-parameter Gumbel–Hougaard copula with LN and WE marginal distributions should be further investigated.

4.5.2. Joint exceedance probability function for $(Q_{max,f}, V_f)$ variables

Figs. 12 and 13 show joint exceedance probability functions for the $(Q_{max,f}, V_f)$ variable for the 2D normal and Gumbel–Hougaard copula-based distributions, respectively. These figures show isolines of joint exceedance probability $P(Q_{max,f} > q_{max,f}, V_f > v_f) = p_0$, where p_0 adopts values from the set $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$. Every point lying on the curve $p = p_0$ defines a rectangle (PABC) with sides parallel to the abscissa and ordinate in such a manner that the probability of random variable $(Q_{max,f}, V_f)$ belonging to this rectangle is p_0. For the purpose of analysis, the quantile with a probability of 0.99 was determined for both random variables $Q_{max,f}$ and V_f respectively:$Q^{0.99}_{max,f} = 852.702\,[m^3\,s^{-1}]$

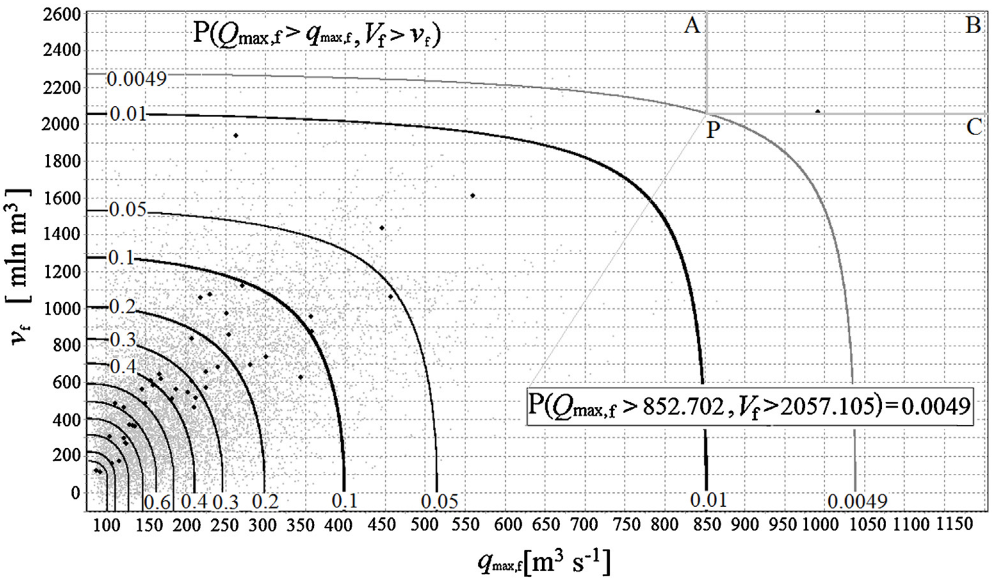

Fig. 13. Exceedance probability function P($Q_{max,f} > q_{max,f}, V_f > v_f$) for 2D distribution of ($Q_{max,f}$, V_f) variables built on the Gumbel–Hougaard copula.

and $V_f^{0.99}$ = 2.057 [10^9 m^3]. The rectangle APBO for p = 0.0038 was defined in Fig. 12, and a 2D normal distribution of the joint exceedance probability $P(Q_{max,f} > q_{max,f}, V_f > v_f)$ was derived from Eq. (2). The values of the random variables $Q_{max,f}$ and V_f were 852.702 [m^3 s^{-1}] and 2057.105 [10^6 m^3], respectively, such that $P(Q_{max,f} > 852.702, V_f > 2057.105) = 0.0038$. In Fig. 12, the small grey points represent generated realizations of the ($Q_{max,f}$, V_f) random variable with normal marginal distributions, whereas the larger black points represent observed historical realizations of this variable.

The point P lying on the curve for p = 0.0049 (Fig. 13) defines the rectangle PABC in such a manner that the probability of random variable ($Q_{max,f}$, V_f) belonging to this rectangle (excluding sides AP and PC) is p = 0.0049. In this case, the joint exceedance probability $P(Q_{max,f} > q_{max,f}, V_f > v_f)$ for $q_{max,f}$ = 852.702 [m^3 s^{-1}] and v_f = 2057.105 [10^6 m^3] was equal to 0.0049. In this case, the small grey points represented generated realizations of the ($Q_{max,f}$, V_f) random variable with log-normal and Weibull marginal distributions.

The K–S, A–D, IA–D, K and L^2 distance tests were employed in the selection of the 2D distribution of the ($Q_{max,f}$, V_f) variables. Assuming an empirical cumulative distribution almost always coincides with a theoretical cumulative distribution, the minimal distance between both cumulative distributions should indicate the best fit of a theoretical distribution to observed data. The two statistical tests, A–D and IA–D, which focus on the distributions' tail regions recognized the 2D distribution generated by the Gumbel–Hougaard copula as being the best fit for the ($Q_{max,f}$, V_f) variables. It should be noted that the IA–D statistic is particularly vulnerable to the occurrence of outliers in the random sample, a property of the IA–D test which is important when a statistical model addresses extreme events (Neyman and Scott, 1971).

4.6. Practical comparisons of Gaussian copula, Gumbel–Hougaard copula and normal distribution flood analysis applications

The joint exceedance probability of snowmelt flood occurrence with peak discharge exceeding 852.702 m^3 s^{-1} and a flood volume exceeding 2057.105 $10^6 \times$ m^3 was equal to 0.0038 when the random variables ($Q_{max,f}$, V_f) yielded a 2D normal probability distribution. However, for the 2D probability distribution generated by a Gumbel–Hougaard copula, the joint exceedance probability was 0.0049. While the difference between these two probabilities is small (0.0011), in the context of water structure design (i.e. dams, water reservoirs, levees, polders, etc.), the exceedence probability is of particular importance. This is due to current national regulations in Poland, suggesting that this probability serves as the basis for the sizing of water control structures (Ozga-Zielinska et al., 2011). For example, if according to a country's particular regulations, the exceedance probability for designing a certain class of water control structures is P = 0.0049, the value $q_{max,f}$ of the variable $Q_{max,f}$ and value v_f of the variable V_f obtained from a 2D normal distribution for this probability P will be less than the equivalent values obtained from the 2D probability distribution generated by a Gumbel–Hougaard copula (Fig. 12). This is illustrated in Fig. 13, where the normally-distributed peak flood discharge $q_{max,f}$ = 852.702 [m^3 s^{-1}] for an exceedance probability of P = 0.0038 is associated with a flood volume v_f, which is less than that obtained from a Gumbel–Hougaard copula-generated probability distribution for the peak discharge. The same situation occurs for a normally-distributed v_f = 2057.105 [$10^6 \times$ m^3] with an exceedance probability P = 0.0038, in the case that the value of peak discharge $q_{max,f}$ is less than the value of $q_{max,f}$ obtained from a Gumbel–Hougaard copula-generated probability distribution for the flood volume. Conversely, the exceedance probability P obtained for the Gumbel–Hougaard copula-generated distribution is greater than that obtained from a normal distribution, indicating that a flood with quantitative character-

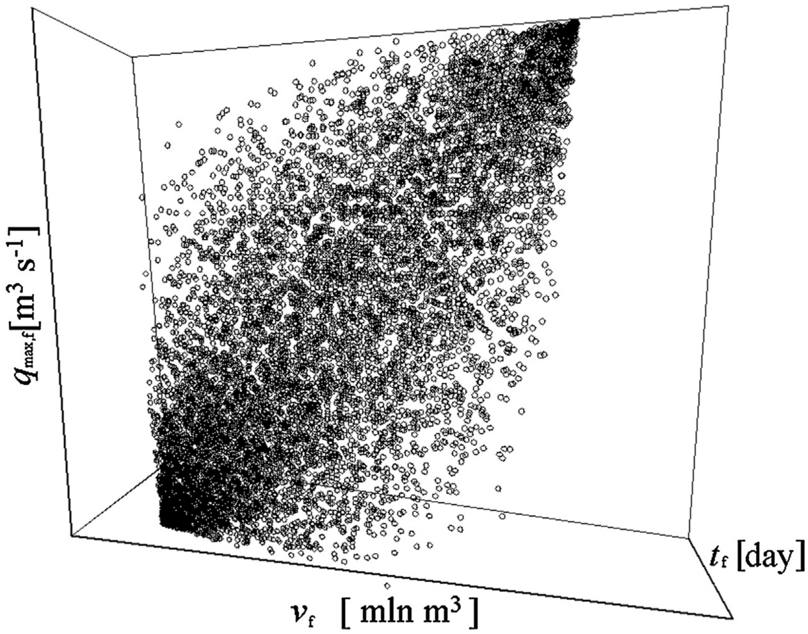

Fig. 14. Chart of generalized realization of 3D variable ($Q_{max,f}$, V_f, T_f) built on the Gumbel–Hougaard 3D copula.

istics of $q_{max,f}$ = 852.702 [m^3 s^{-1}] and v_f = 2057.105 [$10^6 \times$ m^3] would be observed more frequently according to a statistical model estimated by the Gumbel–Hougaard copula method. The consequence of this can be seen when the risk of an event occurrence is assessed. Taking into account the classical approach to risk assessment, the risk R can be computed as $R = P \times L$, where P is the exceedance probability and L is losses (in terms of the number of human lives or economic value). Thus, according to a statistical method estimated by the Gumbel–Hougaard copula method, when it comes to designing hazard zones on the basis of stable (fixed) losses, the potential risk of such an event is greater than that derived from the model of a normal distribution (Ozga-Zielinski, 2015).

Based on the results of this study, it should be noted that, in terms of operational hydrology (snowmelt flood condition, rescue operations in operational mode) for instant quantitative assessment of potential hydrological risks (flood hazard), a sufficiently accurate method is to use the bivariate normal distribution. In the case of activities related to mitigating flood hazard, reducing the effects of floods, recovering from flood events, protecting property from future damage and mapping flood-prone areas, bivariate distributions with parameters estimated by the Gumbel–Hougaard copula method should be used. This should be done with full observance of procedures related to the following activities: preprocessing input data (homogeneity and statistical analysis), selecting optimal marginal distributions and choosing methods of estimating parameters of marginals and copula functions, as well as qualitatively and quantitatively assessing the resulting 2D snowmelt flood model.

For a more complete quantitative description of snowmelt floods, a 3D variable that can be supplemented by a snowmelt flood duration T (e.g. expressed in days) should be applied. In this way, the design hydrograph characteristics of a 3D phenomenon would be composed of maximum peak discharge $Q_{max,f}$, flood volume V_f and duration T_f. While the generation of a 3D variable is not complicated (Fig. 14), the mathematics involved in conducting the probability analysis may lead to difficulties. For example, problems may arise when obtaining the density of a 3D copula that is calculated by its 3rd partial derivate with respect to 3 variables. Therefore, in practice, a simplified solution is based on regression analysis, bivariate conditional distributions, bivariate joint distributions and Kendall distribution functions (De Michele et al., 2007; Evin and Favre, 2008; Gräler et al., 2013).

The most popular solution is a 3D vine copula joining, in our case, the three marginal variables: $Q_{max,f}$, V_f and T_f. The basic idea of a vine copula is to construct high-dimensional copulas based on conditional bivariate copulas (Vernieuwe et al., 2015). Generally, to construct a family of m-variate distributions, two m-1 dimensional marginals having m-2 variables in common are needed (De Michele et al., 2007). The complete density function $c\,(Q_{max,f}, V_f, T_f)$ of a 3D copula is the following:

$$c_{Q_{max,f}V_fT_f}\left(q_{max,f}, v_f, t_f\right) = c_{Q_{max,f}T_f|V_f}\left(F_{Q_{max,f}|V_f}\left(q_{max,f}|v_f\right), F_{T_f|V_f}\left(t_f|v_f\right)\right) \\ c_{Q_{max,f}V_f}(q_{max,f}, v_f) \cdot c_{v_fT_f}\left(v_f, t_f\right) \tag{36}$$

5. Summary and conclusions

With a likely increase in extreme events (e.g. floods) in certain areas in the future due to climate change (Adamowski et al., 2009; Adamowski et al., 2010; Nalley et al., 2012; Nalley et al., 2013; Haidary et al., 2013; Pingale et al., 2014; Araghi et al., 2015), it will become ever more important to protect the public and their property from dangerous natural extreme hydrological events (e.g. large floods); to do so, hydrologists must be able to accurately calculate the probability of occurrence and duration of floods with peak flows or volumes exceeding a certain threshold. This requires the application of a mathematical description of flood occurrence through the use of statistical models and probability distributions. Ultimately, the choice of model must take into consideration the results of the calculated probability $P(Q_{\mathrm{max,f}} > q_{\mathrm{max,f}}, V_{\mathrm{f}} > v_{\mathrm{f}})$ with the smallest error, suggesting that particular attention be placed on the tails (e.g. extreme events) of the models.

This research compared the use of three different models—a 2D normal probability distribution, the elliptical Gaussian copula and the Archimedean 1-parameter Gumbel–Hougaard copula—to the bivariate (2D) analysis of spring snowmelt flood parameters (e.g. peak discharge ($Q_{\mathrm{max,f}}$) and flood volume (V_{f})). A Gaussian copula is not able to accurately model more complex dependence structures (i.e. snowmelt flood described by the random variables ($Q_{\mathrm{max,f}}$, V_{f})); dependence on extreme values relies on marginal values. The Gaussian copula is symmetric and the lower and – more importantly – upper tail dependence coefficients are equal to zero. This property of the Gaussian copula definitely limits its effective use in probabilistic descriptions of complex extreme phenomena, but on the other hand, the simplicity of this function encourages its use. The A–D, IA–D and K tests explicitly pointed to the Gumbel–Hougaard copula-based model as that which best fit the 2D theoretical model for the random variables ($Q_{\mathrm{max,f}}$, V_{f}). As the exceedance probabilities obtained from the Gumbel–Hougaard copula were shown to be greater than those of the normal distribution, the results indicate an increasing risk of higher losses in the river basin if the Gumbel–Hougaard copula model is used for designing flood protection infrastructure (Ozga-Zielinski, 2015).

Copula functions have enabled the separate modelling of marginal distributions and joint distributions. The Gumbel–Hougaard copula allowed for the shortening of the parameter estimation procedure compared to the Gaussian copula. Copula functions have enabled the free choice of marginal distributions in the 2D model, and the consequent shift away from dependence normality or ellipticity models compared to the 2D normal distribution model. Copula functions have also enabled a move away from the linear measure of correlation between studied variables.

The Archimedean copula in the form of Gumbel–Hougaard, coupled with the possibility to choose marginal distributions, enabled us to address several important issues related to the probabilistic description of snowmelt floods. An example of one of these issues is the derogation from the normality related to – among other things – the problem of fat tails, especially important in probabilistic descriptions of extreme natural phenomena such as snowmelt floods. The Gumbel–Hougaard copula made it possible to capture the nonlinear relationship between high values of $Q_{\mathrm{max,f}}$ and relatively lower values of V_{f}, as well as the inverse of events (i.e. the flooding of low maximum flow and large volumes of such floods). The upper tail of the Gumbel–Hougaard copula is heavy (compared to normal), and the relationship between extreme values is described by the upper indices $\tau^{u} = 0.4870$. For the Gaussian copula, $\tau^{u} = 0$, whereas the lower tails in both cases are light (normal). Asymmetries and tail dependence can be taken into account to build models of snowmelt floods, taking advantage of multivariate copula modelling with the separation of marginal and dependence modelling as well as the flexibility of 2D copulas, and the Gumbel–Hougaard copula in particular. To summarize, in the case of snowmelt floods at the Wizna site in Poland, the accurate probabilistic description of extreme events can be obtained by a 2D normal distribution. The advantage of this approach compared to the copula solution is the simplicity of its calculation, as well as better results in terms of the obtained probabilities of exceedance. In light of the research carried out in this study, the recommendation is to use, in the case of bivariate random variables, the 2D normal distribution model in applications where determining design floods is necessary, i.e. designing water management structures like dams, levees, weirs, barriers, barrages, water reservoirs, etc. and for flood risk assessment and management, in particular for designing flood hazard and flood risk zones.

Conflicts of interest

There are no conflicts of interest to report for this paper.

Acknowledgements

This research was supported by an NSERC Discovery Grant held by Jan Adamowski.

References

Abidin, N.Z., Adam, M.B., Midi, H., 2012. The goodness-of-fit test for gumbel distribution. A comparative study. Matematika 28 (1), 35–48.
Adamowski, J., Adamowski, K., Bougadis, J., 2010. Influence of trend on short duration design storms. Water Resour. Manag. 24, 401–413.
Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. Hydrol. Proc. 23, 2686–2696.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control AC 19 (6), 716–722, http://dx.doi.org/10.1109/TAC.1974.1100705.
Aksoy, H., 2000. Use of gamma distribution in hydrological analysis. Turk J. Eng. Environ. Sci. 24, 419–428.

Amal, S.H., 2006. Goodness-of-fit for the generalized exponential distribution. Seen 30 April 2014 at http://interstat.statjournals.net/YEAR/2005/articles/0507001.pdf.
Anderson, R.L., 1941. Distribution of the serial correlation coefficient. Ann. Math. Statist. 8 (1), 1–13.
Andrieu, C., De Freitas, N., Doucet, A., Jordan, M.I., 2003. An introduction to MCMC for machine learning. Mach. Learn. 50, 5–43, http://dx.doi.org/10.1023/a:1020281327116.
Araghi, A., Adamowski, J., Nalley, D., Malard, J., 2015. Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. J. Atmos. Res. 11, 52–72.
Bačová Mitková, V., Halmová, D., 2014. Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. J. Hydrol. Hydromechanics 62 (3), 186–196, http://dx.doi.org/10.2478/johh-2014-0026.
Barker, E., Kelsey, J., 2012. Recommendation for Random Number Generation Using Deterministic Random Bit Generators. National Institute of Standards and Technology, SP800-90A, 128 pp. Seen 30 April 2014 at http://csrc.nist.gov/publications/nistpubs/800-90A/SP800-90A.pdf.
Barndorff-Nielsen, O.E., 2007. Normal inverse gaussian distributions and stochastic volatility modelling. Scand. J. Stat. 24 (1), 1–13, http://dx.doi.org/10.1111/1467-9469.00045.
Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. J. Hydrol. 508, 418–429.
Box, G.E.P., Muller, M.E., 1958. A note on the generation of random normal deviates. Ann. Math. Stat. 29 (2), 610–611, http://dx.doi.org/10.1214/aoms/1177706645.
Brzeziński, J., 2010. Application of generalized exponential distribution in seasonal maximum annual flow analysis. In: Więzik, B. (Ed.), Hydrology in Engineering and Water management. Monograph Environmental Engineering Committee PAN, 71-81 (in Polish with English summary).
Bulletin 17B, 1982, Guidelines for determining flood flow frequency. Bulletin 17B of the Hydrology Subcommittee, Revised September 1981, Editorial Corrections March 1982, Interagency Advisory Committee on Water Data, US Department of the Interior, Geological Survey.
Butler, C., Adamowski, J., 2015. Empowering marginalized communities in water resources management: addressing inequitable practices in participatory model building. J. Environ. Manag. 153, 153–162.
Chapra, S., Canale, R., 2006. Numerical Methods for Engineers, 5th ed. McGraw-Hill Science/Engineering/Math, pp. 960.
Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Method in Finance. John Wiley & Sons, Ltd., pp. 293.
Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Am. Statistician 49 (4), 327–335.
Choroś, B., Ibragimov, R., Permiakova, E., 2010. Copula Estimation. In: Jaworski, P., Darante, F., Karl Härdle, W., Rychlik, T. (Eds.), Copula Theory and Its Applications. Lectures Notes in Statistics, pp. 77–91.
Chowdhary, H., Escobar, L.A., Singh, V.P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data, Hydrol. Res. 42 (2–3), 193–216, http://dx.doi.org/10.2166/nh.2011.065.
Ciupak, M., 2004. Statistical models for forecast of snow-melt flood Methods of sets of predictors and predictands analysis. Rep. Inst. Meteorol. Water Manag. 27 (3–4) (in Polish with English summary).
Ciupak, M., 2011. Multivariate storm surges analysis. The study a capabilities of Copula Theory application, Reports of Institute of Meteorology and Water Management, V (LV) (1–2), 3–23 (in Polish with English summary).
Ciupak, M., 2013. Selection of extreme value probability distribution on example of storm surges and backwater on the Slupia River analysis. In: Weglarczyk, S. (Ed.), Problems of extreme floods estimation in gauged and unguaged catchments. Monography of Polish Water Management Committee of Polish Academy of Sciences, 35, pp. 57–71 (in Polish with English summary).
Dahmen, E.R., Hall, M.J., 1990. Screening of hydrological data: tests for stationarity and relative consistency. ILRI Publication No. 49.
Danaher, P.J., Smith, M.S., 2009. Modeling Multivariate Distribution Using Copulas: Alications in Marketing. Melbourne Business School, University of Melbourne, pp. 47.
D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-fit Techniques. Marcel Dekker, New York, pp. 560.
De Michele, C., Salvadori, G., 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. J. Geophys. Res. 108, http://dx.doi.org/10.1029/2002JD002534.
De Michele, C., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coastal Eng. 54, 734–751, http://dx.doi.org/10.1016/j.coastaleng.2007.05.007.
Directive, 2007/60/EC of the European Parliament and of the Council on the assessment and management of flood risk, Official Journal of the European Union, L 288, vol. 50 from 6.11.2007.
Domino, K., Błachowicz, T., Ciupak, M., 2014. The use of copula functions for predictive analysis of correlations between extreme storm tides. Phys. A: Stat. Mech. Appl. 413 (1), 489–497, http://dx.doi.org/10.1016/j.physa.2014.07.020.
Evin, G., Favre, A.C., 2008. A new rainfall model based on the Neyman-Scott process using cubic copulas. Water Resour. Res. 44, W03433, http://dx.doi.org/10.1029/2007WR006054.
Favre, A.-C., El Adlouni, S., Perreault, Thiémonge, N., Bobeé, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resour. J., 40, http://dx.doi.org/10.1029/2003WR002456.
Ferro, V., Porto, P., 2006. Flood frequency analysis for sicily, Italy. J. Hydrol. Eng. 11 (2), 110–122, http://dx.doi.org/10.1061/(ASCE) 1084-0699(2006)11:2(110).
Ganguli, P., Reddy, M.J., 2013. Probabilistic assessment of flood risks using trivariate copulas. Theor. Appl. Climatol. 111 (1–2), 341–360, http://dx.doi.org/10.1007/s00704-012-0664-4.
Garcia, A., Gençay, R., 2007. Managing adverse dependence for portofolios of collateral in financial infrastructures. Bank of Canada Working Paper 2007–25, 25 pp. Seen 30 April 2014 at http://www.bankofcanada.ca/wp-content/uploads/2010/03/wp07-25.pdf.
Genest, C., Favre, A.C., 2007. Everythink you always wanted to know about copula modeling but were afraid to asl. J. Hydrol. Eng. 12, 347–368, http://dx.doi.org/10.1061/(ASCE) 1084-0699(2007) 12:4(347).
Genest, C., Favre, A.C., Béliveau, J., Jacques, C., 2007. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. Water Resour. Res. 43 (9), http://dx.doi.org/10.1029/2006WR005275.
Genest, C., Quessy, J.F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the integral probability transformation. Scand. J. Stat. 33 (2), 337–366, http://dx.doi.org/10.1111/j.1467-9469.2006.00470.x.
Genest, C., Rémillard, B., Beaudoin, D., 2009. Godness-of-fit tests for copulas: a review and power study. Insurance Math. Econ. 44 (2), 199–213, http://dx.doi.org/10.1016/j.insmatheco.2007.10.005.
Gower, J.C., 1985. Measures of Similarity, Dissimilarity, and Distance in Encyclopedia of Statistical Sciences, 5. In: Kotz, S., Johnson, N.L., Read, C.B. (Eds.). John Wiley and Sons, New York, pp. 397–405.
Gräler, B., van den Berg, M.J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., Verhoest, N.E.C., 2013. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. Hydrol. Earth Syst. Sci. 17, 1281–1296, http://dx.doi.org/10.5194/hess-17-1281-2013.
Grubbs, F.E., Beck, G., 1972. Extention of sample size and percentage points for significance tests of outlying observations. Technometrics 14 (4), 847–854.
Gupta, R.D., Kundu, D., 1999. Generalized exponential distributions. Aust. N. Z. J. Stat. 41 (2), 173–188.
Gumbel, E.J., 1960. Distributions Des Valeurs Extremes En Plusiers Dimensions, 9. Public Institute of Statistics University, Paris, pp. 171–173.
Gwinn, D.A. Sr., 1993. Modified Anderson-Darling and Cramer-Von Mises Goodness-of-Fit Tests for the Normal Distribution. M.Sc Thesis [AFIT/GOR/ENS/93M-7], Operations Research, Air Force Institute of Technology, Air University, Wright-Patterson AFB, OH. Seen 30 April 2014 at http://www.dtic.mil/dtic/tr/fulltext/u2/a262554.pdf.

Haidary, A., Amiri, B.J., Adamowski, J., Fohrer, N., Nakane, K., 2013. Assessing the impacts of four land use types on the water quality of wetlands in Japan. Water Resour. Manag. 27, 2217–2229.
Halbe, J., Adamowski, J., Bennett, E., Pahl-Wostl, C., Farahbakhsh, K., 2014. Functional organization analysis for the design of sustainable engineering systems. Ecol. Eng. 73, 80–91.
Halbe, J., Pahl-Wostl, C., Sendzimir, J., Adamowski, J., 2013. Towards adaptive and integrated management paradigms to meet the challenges of water governance. Water Sci. Technol.: Water Supply 67, 2651–2660.
Hougaard, P., 1986. A class of multivariate failure time distributions. Biometrika 73 (3), 671–678, http://dx.doi.org/10.1093/biomet/73.3.671.
Hougaard, P., 2000. Analysis of Multivariate Survival Data. Springer-Verlag, New York-Berlin-Heidelberg, pp. 542.
Inam, A., Adamowski, J., Halbe, J., Prasher, S., 2015. Using causal loop diagrams for the initialization of stakeholder engagement in soil salinity management in agricultural watersheds in developing countries: a case study in the Rechna Doab watershed. Pak. J. Environ. Manag. 152, 251–267.
Jeong, D.I.l., Sushama, L., Khaliq, M.N., Roy, R., 2013. A copula-based multivariate analysis of Canadian RCM projected changes to flood characteristics for northeastern Canada. Clim. Dyn., http://dx.doi.org/10.1007/s00382-013-1851-4.
Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall London, pp. 424.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, 1. Wiley-Interscience, Wiley Series in Probability and Statistics, pp. 761.
Karmakar, S., Simonovic, S.P., 2009. Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. J. Flood Risk Manag. 2 (1), 32–44, http://dx.doi.org/10.1111/j.1753-318X.2009.01020.x.
Klein, B., Schumann, A.H., Pahlow, M., 2011. Copulas-New Risk Assessment Methodology for Dam Safety. In: Schumann, A.H. (Ed.), Flood Risk Assessment and Management. How to Specify Hydrological Loads, Their Consequences and Uncertainties. Springer Science & Business Media, 149-187, http://dx.doi.org/10.1007/978-90-481-9917-4.
Klonecki, W., 1999. Statistics for Engineers. PWN, Warsaw, pp. 283 (in Polish).
Kolinjivadi, V., Adamowski, J., Kosoy, N., 2014. Recasting payments for ecosystem services (PES) in water resource management: A novel institutional approach. Ecosyst. Serv. 10, 144–154.
Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Bivariate and Trivariate Normal Distribution. In: Continuous Multivariate Distributions, Models and Applications, 2nd ed. Wiley, New York, pp. 251–348.
Kotz, S., Kozubowski, T.J., Podgórski, K., 2001. The Laplace Distribution and Generalizations. In: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Birkhäuser, Boston, pp. 351.
Koziol, J.A., 2008. A weighted Kuiper statisti for goodness of fit. Stat. Neerlandica 50 (3), 394–403, http://dx.doi.org/10.1111/j.1467-9574.1996.tb01505.x.
Krstanovic, P.F., Singh, V.P., 1987. A multivariate stochastic flood analysis using entropy. In: Reidel, D. (Ed.), Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Hydrologic Frequency Modelling. Dordrecht The Netherlands, 515-539, http://dx.doi.org/10.1007/978-94-009-3953-0_37.
Kuchment, L.S., Demidov, V.N., 2013. On the application of copula theory for determination of probabilistic characteristics of springflood. Russian Meteorol. Hydrol. 38 (4), 263–271, http://dx.doi.org/10.3103/S1068373913040080.
Lee, T., Modarres, R., Ouarda, T.B.M.J., 2013. Data-based analysis of bivariate copula tail dependence for drought duration and severity. Hydrol. Processes 27 (10), 1454–1463, http://dx.doi.org/10.1002/hyp.9233.
Li, T.Y., Guo, S.L., Liu, Z.J., Li, L.P., Hong, X.J., 2014. Design flood estimation based on bivariate joint distribution of flood peak and volume. J. Hydraul. Eng. 45 (3), 269–276, http://dx.doi.org/10.13243/j.cnki.slxb.2014.03.003.
Liao, M., Shimokawa, T., 1999. A new goodness-of-fit test for Type-I extreme value and 2-parameter Weibull distributions with estimated parameters. J. Stat. Comput. Simul. 64 (1), 23–48, http://dx.doi.org/10.1080/00949659908811965.
Matúš, R., 2009. The modelling of hydrological joint events on the Morava River using aggregation operators. Slovak J. Civil Eng. 17 (3), 9–15.
McLeish, D.L., Small, G., 1988. The theory and applications of statistical infenece functions. Lect. Notes Stat., 44, http://dx.doi.org/10.1007/978-1-4612-3872-0.
Mendez, B.V.M., de Melo, E.F.L., Nelse, R.B., 2007. Copulas and applications robust fits for copula models. Commun. Stat.-Simul. Comput. 36 (5), 997–1017, http://dx.doi.org/10.1080/03610910701539708.
Myung, J., 2003. Tutorial on maximum likelihood estimation. J. Math. Psychol. 47 (1), 90–100, http://dx.doi.org/10.1016/S0022-2496(02) 00028-7.
Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). J. Hydrol. 475, 204–228.
Nalley, D., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2013. Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform. J. Atmos. Res. 132/133, 375–398.
Nelsen, R.B., 2006. An Introduction to Copulas, 2nd ed. Springer-Verlag, New York, pp. 272.
Neyman, J., Scott, E., 1971. Outlier proneness of phenomena and of related distributions. In: Rustagi, J.S. (Ed.), Optimizing Methods in Statistics. Academic Press, New York, pp. 413–430.
Nourani, V., Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. J. Hydrol. 514, 358–377.
Ozga-Zielinska, M., Brzeziński, J., 1997. Applied Hydrology. PWN, Warsaw, pp. 326 (in Polish).
Ozga-Zielinska, M., Kupczyk, E., Ozga-Zielinski, B., Suligowski, R., Brzezinski, J., Niebala, J., 2011. River-flooding potential in terms of water structures safety and flooding hazard. In: Introduction to Methodology. Institute of Meteorology and Water Management—National Research Institute, Warsaw, pp. 89.
Ozga-Zielinska, M., Ozga-Zielinski, B., Brzeziński, J., 2005. Guidelines for flood frequency analysis. In: Long Measurement Series of River Discharge. Institute of Meteorology and Water Management, Warsaw, pp. 44 (WMO HOMS Component I81.3.01).
Ozga-Zielinski, B., 1999. Methods of hydrological series nonhomogeneity analysis. Rep. Inst. Meteorol.Water Manag. 2, 44 (in Polish with English summary).
Ozga-Zielinski, B., 2015. Safety and Reliability of Hydrologic Systems. In: Scientific Works of the Warsaw University of Technology—Environmental Engineering, No. 69. Publishing House of the Warsaw University of Technology, pp. 125 (in Polish with English summary).
Pilon, P.J., Condie, R., Harvey, K.D., 1985. Consolidated Frequency Analysis Package CFA User Manual for Version—DEC PRO SERIES. Water Resources Branch, Inland Waters Directorate Enviroment Canada, Ottawa.
Pingale, S., Khare, D., Jat, M., Adamowski, J., 2014. Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan. India J. Atmos. Res. 138, 73–90.
Poulin, A., Huard, D., Favre, A.-C., Pugin, S., 2007. Importance of tail dependence in bivariate frequency analysis. J. Hydraul. Eng. 12 (4), http://dx.doi.org/10.1061/(ASCE) 1084-0699(2007)12:4(394).
Renard, B., Lang, M., 2007. Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. Adv. Water Resour. 30, 897–912, http://dx.doi.org/10.1016/j.advwatres.2006.08.001.
Rose, C., Smith, M.D., 2002. The Bivariate Normal. In: Mathematical Statistics with Mathematica. Springer-Verlag, New York, pp. 216–226.
Saad, C., El Adlouni, S., St-Hilaire, A., Gachon, P., 2014. A nested multivariate copula approach to hydrometeorological simulations of spring floods: the case of the Richelieu River (Québec Canada) record flood. Stoch. Environ. Res. Risk Assess. 29 (1), 275–294, http://dx.doi.org/10.1007/s00477-014-0971-7.
Sadri, S., Burn, D.H., 2014. Copula-based pooled frequency analysis of droughts in the canadian prairies. J. Hydrol. Eng. 19 (2), 277–289, http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000603.
Sagias, N.C., Karagiannidis, G.K., 2005. Gaussian class multivariate Weibull distributions: theory and applications in fading channels. Inf. Theory IEEE Trans. 51 (10), 3608–3619, http://dx.doi.org/10.1109/TIT.2005.855598.
Salvadori, G., De Michele, C., 2007. On the use of copulas in hydrology. theory and practice. J. Hydrol. Eng. 12, 369–380, http://dx.doi.org/10.1061/(ASCE) 1084-0699(2007)12:4(369).

Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R., 2007. Bivariate analysis via Copulas. In: Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R. (Eds.), Extreme in Nature. An Approach Using Copulas, 56. Springer Netherlands, Water Science and Technology Library, pp. 131–175, http://dx.doi.org/10.1007/1-4020-4415-1.
Singh, V.P., Wang, S.X., 2005. Frequency analysis of nonidentically distributed hydrologic flood data. J. Hydrol. 307 (1–4), 175–195, http://dx.doi.org/10.1016/j.jhydrol.2004.10.029.
Sklar, A., 1959. Fonctions De Répartition Á Dimensions Et Leurs Marges, 8. Public Institute of Statistics University, Paris, pp. 229–231.
Sneyers, R., 1990. On the statistical analysis of series of observations. TN No.143, WMO, No. 415.
Song, S., Singh, V.P., 2010. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch. Environ. Res. Risk Assess. 24 (3), 425–444, http://dx.doi.org/10.1007/s00477-009-0331-1.
Sraj, M., Bezak, N., Brilly, M., 2014. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. Hydrol. Process., http://dx.doi.org/10.1002/hyp.10145, in press.
Stedinger, J.R., Griffis, V.W., 2008. Flood frequency analysis in the united states: time to Update. J. Hydrol. Eng. 13 (4), 199–204, http://dx.doi.org/10.1061/(ASCE) 1084-0699(2008) 13:4(199).
Stephens, M.A., 1969. A goodness-of-fit statistic for the circle, with some comparisons. Biometrika 56 (1), 161–168.
Straith, D., Adamowski, J., Reilly, K., 2014. Exploring the attributes, strategies and contextual knowledge of champions of change in the Canadian water sector. Can. Water Resour. J. 39 (3), 255–269.
Strupczewski, W.G., 1967. Transformation probability distributions of hydrological and meteorological variables into normal distribution. Rep. Inst. Meteorol. Water Manag. 2 (in Polish with English summary).
Vernieuwe, H., Vandenberghe, S., De Baets, B., Verhoest, N.E.C., 2015. A continuous rainfall model based on vine copulas. Hydrol. Earth Syst. Sci. 19, 2685–2699, http://dx.doi.org/10.5194/hess-19-2685-2015.
Węglarczyk, S., 1993. Goodness-of-fit test for distributions with estimated parameters. Rev. Geophys. 65 (3–4) (in Polish with English summary).
Węglarczyk, S., 2010. Statistics in environmental engineering. Publishing House of the Cracow University of Technology, pp. 375 (in Polish).
Wong, G., Lambert, M.F., Leonard, M., Metcalfe, A.V., 2010. Drought analysis using trivariate copulas conditional on climatic states. J. Hydrol. Eng. 15 (2), 129–141, http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000169.
Yue, S., 1999. Applying bivariate normal distribution to flood frequency analysis. Water Int. 24 (3), 248–254, http://dx.doi.org/10.1080/02508069908692168.
Zhang, L., 2005. Multivariate hydrological frequency analysis and risk mapping. Ph.D. Thesis, Department of Civil and Environmental Engineering, Louisiana State University. Seen 15 September 2013 at http://etd.lsu.edu/docs/available/etd-04122005-214053/unre-stricted/Zhang-dis.pdf.
Zhang, L., Singh, V., 2006. Bivariate flood frequency analysis using the copula method. J. Hydrol. Eng. 11 (2), 150–164, http://dx.doi.org/10.1061/(ASCE) 1084-0699(2006) 11:2(150).
Zhang, L., Singh, V., 2007. Bivariate rainfall frequency distributions using Archimedean copulas. J. Hydrol. 332 (1–2), 93–109, http://dx.doi.org/10.1016/j.jhydrol.2006.06.033.